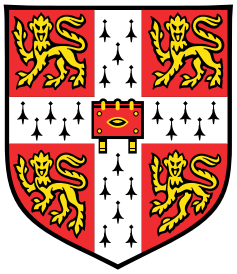
Monitoring trace levels of ctDNA using integration of variant reads



Jonathan C. M. Wan

Supervisor: Dr Nitzan Rosenfeld

Cancer Research UK Cambridge Institute

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Trinity College May 2019

Abstract

In patients with early-stage cancer, ctDNA detection rates can be low due to the presence of few or no copies of any individual mutation in each sample. Sensitivity may be increased by collecting larger plasma volumes, but this is not feasible in practice. Although cancers typically have thousands of mutations in their genome, previous analyses measured only individual or up to 32 tumour-specific mutations in plasma.

Here, we demonstrate that sensitivity can be greatly enhanced for any given input DNA mass by analysing a large number of mutations via sequencing. We sequenced in plasma 10^2-10^4 mutated loci per patient, using custom capture panels, whole exome (WES) or whole genome sequencing (WGS). We developed a method for INtegration of VAriant Reads (INVAR) that aggregates reads carrying tumour mutations across multiple mutant loci, and assigns confidence to error-suppressed reads based on mutation context, fragment length and tumour representation. This workflow combines a number of concepts in a novel approach in order to quantify ctDNA with maximal sensitivity.

We applied INVAR to plasma sequencing data from 45 patients with stage II-IV melanoma and 26 healthy individuals. ctDNA was detected to 1 mutant molecule per million, and tumour volumes of ~1cm^3. We show that this algorithm is applicable across targeted and untargeted sequencing methods. In patients with stage II-III melanoma who relapsed after resection, ctDNA was detected within 6 months post-surgery and prior to relapse in 50% of cases, compared to 16% in a similar cohort analysed with digital PCR. In addition, INVAR may enhance detection of ctDNA in samples with limited input or sequencing coverage by

aggregating signal across a large number of mutations. Using low-depth WGS (0.6x), ctDNA was detected to 1 mutant per 10,000 molecules. Given that 60 genome copies of cfDNA may be obtained from 1 drop of blood (50-75μL), we suggest that INVAR may enable cancer monitoring from limited samples volumes.

As tumour sequencing becomes more widespread allowing identification of a large number of mutations per patient, this method has potential to quantify ctDNA with enhanced sensitivity, and to enable routine cancer monitoring using low-depth sequencing, potentially from low-volume blood samples that might be self-collected.

Lay Abstract

Cancer cells release their contents into the bloodstream when they die. This includes mutated DNA fragments from the cell nucleus, which can be called circulating tumour DNA (ctDNA). Therefore, cancer can be detected from blood tests by identification of ctDNA through methods such as DNA sequencing. The fraction of DNA fragments that are mutated in a patient's blood corresponds to the extent of cancer in their body. Thus, in patients with low disease burden, for example post-surgery or in response to treatment, there are few mutant DNA molecules in the bloodstream. This results in ctDNA molecules being missed when blood is sampled and tested for specific mutations.

Cancers can have up to 10^2 to 10^4 mutations in their genome, each in different regions of the genome. Each mutation gives rise to a ctDNA fragment, representing an opportunity to detect cancer. DNA sequencing-based approaches have so far targeted up to 40 mutations per patient in order to improve sensitivity for ctDNA compared to methods targeting single mutations. In order to target multiple mutations efficiently, These mutations may be identified per patient if tumour sequencing is performed prior to ctDNA analysis, which is possible in the disease monitoring setting, and may allow improved detection or monitoring of cancer.

In this study, we studied patients with early or advanced melanoma skin cancer and identified up to 5,073 tumour mutations per patient, then sequenced patients' blood samples. To achieve high sensitivity detection, we developed a method called 'INtegration of VAriant Reads' (INVAR) that aggregates signal across thousands of mutations, and considers biological and technical features of ctDNA sequencing. This method detects low levels of

ctDNA, down to a single mutant molecule per million circulating DNA molecules. ctDNA was detected in 100% of advanced melanoma patients, and disease was monitored to the limit of CT imaging. Within 6 months post-surgery, this method detected ctDNA in 50% of stage II-III melanoma patients who later relapsed within 5 years. Furthermore, sensitive detection of ctDNA can be achieved from samples where little starting material, or limited (low-depth) sequencing data, is available. We show that this method can monitor ctDNA to 1 mutant molecule per 10,000 using limited input and low-depth WGS. These data could be generated from ~1 genome copy, indicating that cancer monitoring from droplet volumes of blood with high sensitivity may be possible.

As tumour and plasma sequencing becomes increasingly routine, INVAR could enhance detection of residual disease and could enable cancer monitoring using low-depth sequencing, or from limited sample volumes such as blood droplets, which might enable self-collection at home.

Dedicated to my family.

Thank you for always encouraging me to pursue my interests.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Jonathan C. M. Wan

May 2019

Acknowledgements

During this PhD, I have developed greatly both as a scientist, and personally. I am grateful to have been around such excellent scientists in the Rosenfeld lab and at Cancer Research UK, from whom I have learned very much. It has certainly been inspiring to work alongside such motivated and hard-working individuals, who share my aim of improving cancer care for patients.

First, I would like to thank Nitzan Rosenfeld for his excellent support, supervision, and ideas. I feel very privileged to have been able to work with one of the leaders of the ctDNA field; I have been impressed how Nitzan combines a clinician-like eye for translational research questions with the problem-solving approach of a biotechnologist. He offered me excellent support at strategic times, while giving me the academic freedom to creatively problem-solve. I have been consistently impressed by his vision for our work, and am sure the Rosenfeld lab will continue to be world-leading.

During my first year, I was co-supervised by Charlie Massie. During the early stages of my PhD, I needed a higher level of support, and Charlie was generous with his time to support and guide me through this period. I recall thanking him with a Christmas card of that year, writing that he had provided me with the best start to a PhD I could have hoped for. Almost three years on, I reiterate these thanks.

In the lab, I have worked closely with a multidisciplinary team of biologists, computational biologists, mathematicians, and clinicians. I learned my wet lab skills from Suzanne Murphy, Christopher Smith, Florent Mouliere, Katrin Heider, Wendy Cooper, Irena Hude-

cova, Davina Gale and Charlie Massie, all of whom took the time to teach me during my PhD. Through their teaching and support, I have developed a more systematic approach to problem-solving in science. I thank James Morris, Dineika Chandrananda and Francesco Marass for mentoring me through my development of my bioinformatics skills. In addition, I would like to thank the following people for their discussions and help over the years, either clinically or technically: Dee Lynskey, Mareike Thompson, Gahee Park, Keval Patel, Dana Tsui, Jenny Chan, Andrew Gill, Andrea Marshall, Roy Rabbie and Hugh O'Brien. Eyal Fisher joined us in the lab only very briefly, but I would like to thank him for his openness in discussion of mathematics with a non-expert, which I greatly enjoyed.

As part of the MelResist study, I have had the good fortune to collaborate with excellent clinicians and associated staff. I would like to thank Pippa Corrie and Christine Parkinson for their valuable discussions, and in particular for their emphasis on the clinical relevance of our work. I would also like to thank Emily Barker, Cathy Thorbinson and Gemma Young for running the MelResist so well, enabling us to try innovative approaches.

I have worked most closely with Katrin Heider in the lab. We have worked on the INVAR algorithm together, and I am very proud of what we have achieved together. Despite having almost entirely opposite personalities and different problem-solving strategies, we have had a very productive working relationship. This has been the longest and most in-depth team-working experience for me, which has been quite a learning curve - thank you for your patience with me. I am very grateful to have worked so closely with someone so bright and motivated to achieve highly.

On a personal level, I have been supported both by my colleagues, my friends and my family. This PhD has been the most challenging endeavour I have undertaken to date, by virtue of its length, coupled with the nature of science. My colleagues and friends have supported me through all manner of set backs, and I am more resilient as a result and grateful to them. My family have always been there for me, always welcoming me home at every

opportunity, and providing me with overwhelming support for each of my ambitions - they have always made me feel that anything is achievable so long as you work hard. I am very lucky to have such a positive and supportive network around me.

I applied to the Cambridge MB/PhD programme with the ultimate aim of becoming a clinician-scientist. I would like to thank Dr Diana Wood at the Clinical School and Prof. Rebecca Fitgerald at Trinity for providing me with this opportunity to pursue my long-standing career ambition. Although the PhD has been challenging, it was hugely enjoyable, and my desire to pursue a clinical-academic career has been reinforced. I look forward to returning to medical school to develop the 'clinician' part of clinician-scientist, with a view to carrying out both research and clinical practice in parallel in future, so that I may apply myself to help cancer patients and their families through both my clinical skills and my academic creativity.

Postscript: I am very grateful to Dr Moritz Gerstung and Dr Nicholas Turner for being my thesis examiners. They were challenging but fair - I greatly enjoyed the process. I thank them both for their time.

Table of contents

List of figures

List of tables

Nomenclature

Acronyms / Abbreviations

AF	Allele fraction
AVAST-M	Adjuvant bevacizumab in patients with melanoma at high risk of recurrence
bp	Base pairs
CAPP-Seq	CAncer Personalized Profiling by Deep Sequencing
cfDNA	Cell-free DNA
CNA	Copy-number alteration
CRUK	Cancer Research UK
CTC	Circulating Tumour Cell
CT	Computed Tomography
ctDNA	Circulating tumour DNA
dPCR	Digital Polymerase Chain Reaction
EDTA	Ethylenediaminetetraacetic acid
eTAm-Seq	Enhanced Tagged-Amplicon Sequencing

FFPE	Formalin-Fixed Paraffin-Embedded
H&E	Haemotoxylin and Eosin
hGA	Haploid genomes analysed
INVAR	Integration of Variant Reads
IR	Informative reads
irRECIST	Immune-related Response Evaluation Criteria In Solid Tumors
MELR	MelResist
MRD	Minimal Residual Disease
NGS	Next-Generation Sequencing
NIPT	Non-invasive Prenatal Testing
NSCLC	Non-small cell lung cancer
PCR	Polymerase Chain Reaction
PFS	Progression-Free Survival
RECIST	Response Evaluation Criteria In Solid Tumours
ROC	Receiver Operating Characteristic
sWGS	Shallow Whole-Genome Sequencing
TAm-Seq	Tagged-Amplicon Sequencing
TAPAS	Tailored-Panel Sequencing
UV	Ultra-violet
WGS	Whole-Genome Sequencing

Chapter 1

Liquid biopsies come of age: towards implementation of circulating tumour DNA

1.1 Attribution

This chapter is adapted from a review article which we published in 2017:

Wan, JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, Pacey S, Baird R & Rosenfeld N (2017). Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nature Reviews Cancer*, 17(4), 223-238.

In this thesis, I have adapted the manuscript to expand upon Section 1.9 (*Enhancing detection of residual disease*).

For this chapter, I performed the literature review wrote the draft manuscript, under the supervision of Nitzan Rosenfeld and Richard Baird. Following feedback from Nature Reviews Cancer to increase the level of discussion throughout the manuscript, I rewrote

much of the manuscript, synthesising ideas across the cfDNA field. This involved linking ideas between non-invasive prenatal testing and ctDNA, and linking potential applications between cfDNA, cfRNA and other circulating nucleic acids. Nitzan oversaw this re-writing, corrected the wording throughout the Review. The published version is attached.

Given the focus of my thesis on the sensitivity of methods for ctDNA, I added Section 1.10 'Enhancing detection of residual disease', which I drafted and Nitzan Rosenfeld approved. This section allowed the inclusion of studies published on ctDNA detection since the time of our Review being written.

1.2 Abstract

Improvements in genomic and molecular methods are expanding the range of potential applications for circulating tumour DNA (ctDNA), both in a research setting and as a 'liquid biopsy' for cancer management. Proof-of-principle studies have demonstrated the translational potential of ctDNA for prognostication, molecular profiling and monitoring. The field is now in an exciting transitional period in which ctDNA analysis is beginning to be applied clinically, although there is still much to learn about the biology of cell-free DNA. This is an opportune time to appraise potential approaches to ctDNA analysis, and to consider their applications in personalised oncology and in cancer research.

1.3 Introduction

The presence of fragments of cell-free nucleic acids in human blood was first described in 1948 by Mandel and Métais [1]. The origins and characteristics of cell-free DNA (cfDNA) were studied intermittently in subsequent decades [2]. In healthy individuals, cfDNA concentration tends to range between 1-10ng/millilitre (ml) in plasma [2, 3], and raised cfDNA levels were first reported in the serum of cancer patients in 1977 [2]. cfDNA concentration can

also be raised by other illnesses or conditions, such as acute trauma [4], cerebral infarction [5], exercise [6], transplantation [7], and infection [8]. Furthermore, the identification of fetal DNA sequences in maternal plasma by Dennis Lo and colleagues in 1997 [9] has led to multiple applications of cfDNA in prenatal medicine including sex determination [10], identification of monogenic disorders [11], and non-invasive prenatal testing (NIPT) for aneuploidies such as Down's Syndrome (trisomy 21), which was first demonstrated in 2007 by Lo et al. [12] and has moved rapidly into widespread clinical use [13, 14].

In 1989, Stroun, Anker et al. identified that at least some cfDNA in the plasma of cancer patients originates from cancer cells [15]. In 1992, Vogelstein, Sidransky and colleagues showed that DNA from urinary sediments (cell pellets) from patients with invasive bladder cancer carried mutations in *TP53*, setting the stage for the use of genomic analysis methods in liquid biopsy applications [16]. *KRAS* mutations were soon found in stool or sputum that matched mutations from colorectal , pancreatic [17], or lung cancers [18, 19]. In 1994, mutated *KRAS* sequences were first reported to be detected in plasma cfDNA of patients with pancreatic cancer using polymerase chain reaction (PCR) with allele-specific primers [20]. For each patient, the *KRAS* mutation found in the plasma was identical to that found in the patient's tumour, thereby confirming that the mutant DNA fragments in plasma were of tumour origin. Mutations in cfDNA are highly specific markers for cancer, which gave rise to the term circulating tumour DNA (ctDNA).

In the following decades, ctDNA was explored as a prognostic or predictive marker [21, 22] and for cancer detection [23]. Such studies confirmed the potential of ctDNA, though the levels of ctDNA in different clinical contexts were not yet accurately defined. These studies nonetheless could demonstrate potential clinical applications, for example detection of *KRAS* mutations in plasma as a potential prognostic factor in colorectal cancer [24]. The introduction of a digital PCR (dPCR) method in 1999 by Vogelstein and Kinzler enabled the accurate identification and absolute quantification of rare mutant fragments

[25]. A modification of this technique using beads in emulsions [26] and flow cytometry allowed the quantification of the mutant allele fraction of cancer mutations in the plasma of patients with different stages of colorectal cancer [27]. Diehl, Diaz and colleagues then showed in 2008 that ctDNA is a highly specific marker of tumour dynamics, and may be able to indicate residual disease [28]. In parallel, allele-specific PCR and other methods were devised and tested for their ability to identify epidermal growth factor receptor (*EGFR*) mutations in serum or plasma of lung cancer patients, following the elucidation of the role of such mutations in predicting response to treatment with molecularly targeted inhibitors.

The development of next generation sequencing-based technologies has facilitated the interrogation of the genome at a broader scale. In 2012, deep sequencing of multiple genes in cfDNA was demonstrated using panels of tagged amplicons, which allowed the identification of mutations directly in the plasma of cancer patients, and monitoring of multiple tumour-specific mutations in a single assay [29]. This method was subsequently applied to monitor ctDNA in a cohort of patients with metastatic breast cancer [30]. Shortly thereafter, whole-genome sequencing (WGS) of plasma cfDNA was first shown to identify tumour-derived chromosomal aberrations [31], focal amplifications [32] and gene rearrangements [33], and hybrid-capture sequencing was introduced as a non-invasive method to analyse the evolving genomic profile of mutations in cancer across the entire exome [34].

1.4 cfDNA and ctDNA biology

cfDNA is thought to be released from cells mostly through apoptosis and necrosis, and possibly also active secretion [35–39]. Outside of the blood circulation, cfDNA has been detected a variety of body fluids including urine [40–43], cerebrospinal fluid (CSF) [44–46], pleural fluid [47] and saliva [48]. Genetic and epigenetic modifications of cfDNA molecules reflect the genome or epigenome of the cell of origin [49–51] (Figure 1.1). Methylation analysis has revealed that the majority of cfDNA in plasma is released from haematopoietic

cells in healthy individuals [50–52]; these have been suggested to be the source of cfDNA release following intense exercise [53]. Observational studies have determined the half-life of cfDNA in the circulation as between 16 minutes and 2.5 hours [28, 44, 54, 55] which allows ctDNA analysis to be considered as a 'real-time' snapshot of disease burden. Other observational studies indicate that cfDNA is cleared from the circulation via nuclease action [55, 56] and renal excretion into the urine [57, 40]. In addition, cfDNA uptake in the liver and spleen, followed by degradation by macrophages, may also contribute [27, 58]. The stability of individual fragments in the circulation may be increased through association with cell membranes, extracellular vesicles or proteins [35].

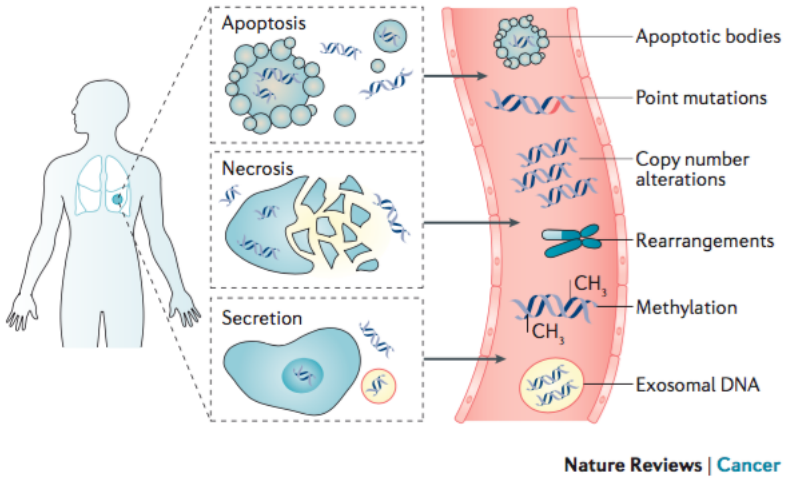


Fig. 1.1 **Origins and spectrum of alterations in cell-free DNA**
Cells release cfDNA through a combination of apoptosis, necrosis, and secretion. cfDNA can arise from cancerous cells but also from cells in the tumour microenvironment, immune cells, or other body organs. In the bloodstream cfDNA may exist as either free, or associated with extracellular vesicles such as exosomes [35]. Multiple classes of genetic and epigenetic alterations can be found in cfDNA. Taken from Wan et al. [59].

Nearly two decades ago, the modal size of cfDNA was determined using gel electrophoresis as ~180 base pairs (bp), indicating that cfDNA was likely to be nucleosome-associated [60]. Sequencing-based approaches have since refined this measurement, by identifying a prominent peak at 166bp [61, 62], corresponding to the length of DNA wrapped around a

nucleosome (~147bp), plus linker DNA associated with histone H1. Fragment size traces of cfDNA show a 10bp ladder pattern [39, 57], ostensibly caused by nucleases cleaving the DNA strand at periodically exposed sites with each turn of the DNA double-helix. The fragmentation patterns of cfDNA differ between plasma and urine [57], potentially contributed to by a higher nuclease activity in urine [63].

ctDNA molecules (i.e. mutant cfDNA fragments) are shorter than non-mutant cfDNA in plasma, demonstrated by PCR [3, 64] and sequencing [51, 65]. Animal xenograft experiments [62, 64–66] provide an elegant means to interrogate ctDNA, since any human DNA sequences must have originated from the tumour xenograft. The modal length of ctDNA fragments has been measured in a rat xenograft model as between 134-144bp [65], though the cause of this shortening is not clear. Shortening of fragments is also observed in fetal cfDNA relative to maternal cfDNA [61], and between non-haematopoetically-derived vs. haematopoietically-derived cfDNA fragments in transplant patients [67, 68]. Differences in nucleosome wrapping or nuclease action between haematopoietic cells, which contribute most to the cfDNA pool, and other tissues may play a role. Long cfDNA fragments (>1,000bp) have been observed in healthy individuals using long-read sequencing techniques [69], and may be released into the circulation in association with exosomes [36, 37], or by tumour cells via necrosis [38]. Current extraction methods often poorly recover these long fragments [6, 70]. Commonly used library preparation methods introduce further biases. Single-stranded DNA (ssDNA) library preparation [71] can recover DNA fragments with damaged ends, and when applied to cfDNA [49, 72] uncovered a large proportion of fragments shorter than 100bp. Diverse extraction and sequencing methods can therefore yield complementary data; combining those with histological analysis of corresponding tissue samples could provide new insights into the biological determinants of cfDNA fragmentation, and the biological origins of cfDNA.

1.4.1 Physiological and pathological considerations

cfDNA has been proposed as a ligand for Toll-like receptor 9 (TLR9) [73, 74], which is a sensor of exogenous DNA fragments that is found primarily in tissues rich in immune cells. In mice, obesity-related adipocyte degeneration was shown to release cfDNA, which contributed to macrophage accumulation via TLR9 activation and led to adipose tissue inflammation and insulin resistance [75]. Another study has suggested that cfDNA may inhibit pro-apoptotic caspases via TLR9-dependent signalling [74] which could imply a potential immunomodulatory role for cfDNA.

In vitro experiments suggest that cfDNA may be internalized by cells [75, 76, 35], which raises the possibility that cfDNA molecules could mediate the horizontal transfer of genes or DNA. One report showed that NIH-3T3 mouse cells that were in contact with samples of plasma from patients with *KRAS*-mutant colorectal cancers underwent *in vitro* transformation, despite being separated to avoid tumour cell contamination [77]. Another study demonstrated the integration of ctDNA into the nuclear DNA of recipient cells and suggested that this may occur through non-homologous end-joining [76]. A similar phenomenon has been observed for mitochondrial DNA [78]. The evidence for this potential phenomenon of *genometastasis* is, at present, limited to a small number of studies and report and so further investigation is required before conclusions can be drawn. In sum, it is clear that there is a lot to learn about the biology of cfDNA and ctDNA, and improved knowledge could have an important impact on their potential applications in oncology.

1.5 Pre-analytical considerations for ctDNA analysis

In low-burden disease, or certain cancer types, the concentration of ctDNA molecules may be low and any loss of sampled material could reduce the sensitivity of molecular profiling. For

quantitative applications, reproducibility of measurement is essential to achieving a robust result, and so the following pre-analytical factors should be considered:

- Samples should be collected in tubes containing an anticoagulant that is compatible with PCR, with EDTA being preferred. Plasma from heparinised blood leads to inhibition of PCR [79], although some studies have successfully utilised such samples [80].

- It is important that the first centrifugation of the blood is done within a few hours of the blood draw in order to remove blood cells that may lyse and release germline DNA which would dilute ctDNA [52, 81–83]. Tubes containing fixative agents may stabilise cells and prevent lysis for several days at room temperature [82, 84–86], including during shipping [86, 87] .

- Following centrifugation, buffy coat DNA from the same tubes can be used as a source of germline DNA, although this may contain small or trace amounts of ctDNA.

- From a blood draw, plasma is preferred over serum for ctDNA analysis [83]. Serum also contains ctDNA [88], but blood cell lysis during the preparation of serum samples releases DNA from non-cancerous cells, which would dilute any ctDNA signal. Other body fluids or cytological specimens may be used, and may contain a higher amount or concentration of tumour DNA depending on tumour proximity.

- cfDNA extraction may be carried out with affinity-column, magnetic bead, polymer, and phenol-chloroform methods. Different methods show variation in their ability to recover particular fragment sizes [6, 70], which could have implications for ctDNA detection, given the differences in size between cfDNA and ctDNA.

1.6 Approaches for ctDNA analysis

Analysis of ctDNA ranges in scale from single mutations to whole-genome analyses (Figure 1.2, Table 1.1). Appropriately designed assays for individual mutations can achieve high sensitivity using a simple workflow. Allele-specific PCR methods have been applied since the mid-2000s for detection of hot-spot mutations in serum and plasma, and some assays are available as kits that are approved for clinical use [89, 90], but have limited analytical sensitivity. dPCR assays on microfluidic platforms are quantitative and highly sensitive, and are used extensively to quantify ctDNA levels [26–28, 42, 91–94]. Improved detection at selected loci has been demonstrated by methods such as single-base extension or enrichment for mutant alleles by electrophoretic methods [95, 96], nuclease activity or modified PCR [97, 98]. The multiplexing capacity of such assays, that rely on differential binding affinities of mutant and wild type alleles, and for the most part require primers or probes that are specific to a defined mutation or targeted locus, is limited. These are therefore generally suited to investigating a small number of mutations, and are often applied to analysis of cancer hot-spot mutations. If samples need to be split into multiple reactions, this increases the likelihood of sampling error, and may impair the overall performance of an assay for very low copy numbers of mutant DNA.

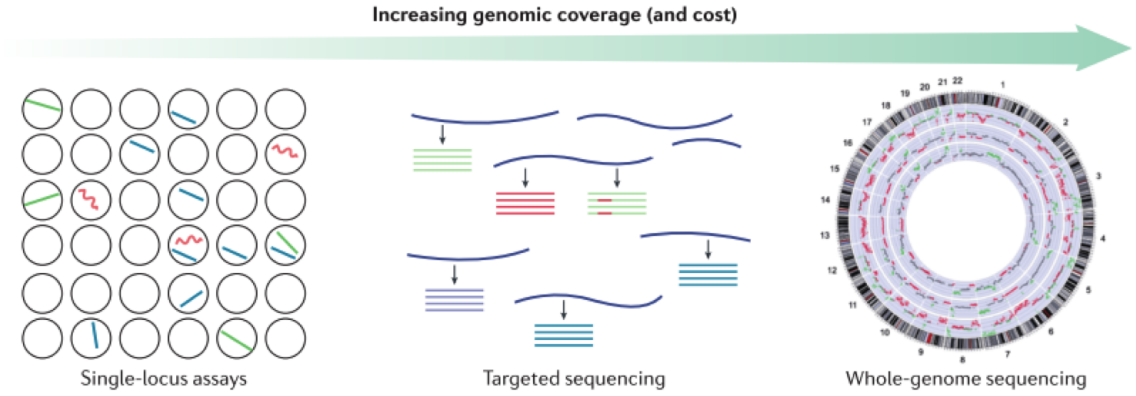


Fig. 1.2 **Scales of analysis of ctDNA**
The analysis of cfDNA can range from the interrogation of individual loci, to analysing the whole genome. Off-the-shelf dPCR assays can achieve high sensitivity with a simple workflow, but are limited by a low multiplexing capability. Targeted sequencing can allow the interrogation of multiple loci with high sensitivity, using methods that suppress background noise [99]. The targeted sequencing image is modified with permission from [29] and the whole genome sequencing image was kindly provided by Dennis Lo, based on data published in ref. [100]. Taken from Wan et al. [59].

Table 1.1 **Comparison and utility of technology platforms for circulating tumour DNA analysis**.
ARMS, amplification-refractory mutation system; BEAMing, beads, emulsion, amplification, and magnetics; CAPP-Seq, cancer personalized profiling by deep sequencing; COLD-PCR, co-amplification at lower denaturation temperature PCR; FAST-SeqS, fast aneuploidy screening test-sequencing system; LINE-1, long interspersed nucleotide element-1. Adapted from [59].

Scale of analysis	Example technologies	Loci interrogated	Indicative limit of detection	Clinical utility
Single-locus or multiplexed assays	Microfluidic or allele-specific PCR: • Digital PCR, • BEAMing Enrichment for mutant alleles: • COLD-PCR, • SCODA	• 1-10 loci • Both ctDNA and cfDNA (IntPlex) 10-100 loci	Varies by method, optimal implementations can reach sensitivity of 0.001%-0.01% or individual mutant copies/ml [27, 101–103]	• Detection and quantification of recurrent hotspot mutations • Monitoring for recurrent resistance mutations
Targeted sequencing approaches	Amplicon-based: • TAm-Seq, • Safe-SeqS Hybrid-capture: • Exome sequencing, • CAPP-Seq, • Digital Sequencing	10kb - 10mb	• <0.01%-0.5% for purpose-built panels [29, 104–107] • 1% for off-the-shelf panels [108, 109] • 5% for exome sequencing [110]	• Profiling gene panels, • Monitoring for *de novo* resistance mutations • Monitoring clonal evolution in response to therapy • Sensitivity for disease burden can be increased by testing multiple loci in parallel
Genome-wide	WGS: • Plasma-Seq, • PARE Amplicon-based: • FAST-SeqS, • mFAST-SeqS	• 3.2gb (whole-genome) • 21.6kb (LINE-1)	• 5-10%	• Identification of structural variants • Stratification of patient samples based on disease burden • Detecting the presence of chromosomal aberrations

In order to interrogate a larger number of loci, targeted sequencing using PCR amplicons or hybrid-capture have been employed [29, 34, 111, 108]. Regions for sequencing may range from individual exons of interest (kilobases), to the entire exome (~50 megabases). Current off-the-shelf panels for gene sequencing can detect mutations with an allele fraction greater than 1% [108, 109]. Amplicon-based assays that have been optimised for the purpose of ctDNA analysis use bespoke analytical and statistical processes, thereby targeting dozens to hundreds of amplicons across multiple kilobases [29, 112, 104]. Hybrid-capture-based approaches can increase the genomic region studied to dozens or hundreds of kilobases [111, 105–107]. The sensitivity for ctDNA detection can be further enhanced, even with limited amounts of input material, by using multiplexed patient-specific panels in combination with targeted sequencing methods [29, 106, 113].

Shallow WGS (sWGS) has been employed to detect fetal aneuploidies [32], and it can also be used to detect cancer-specific copy number alterations [31, 32, 114]. Amplifications and deletions may be identified through low-depth (~0.1x coverage) sequencing of the whole genome, with comparison of the relative number of sequencing reads between equally sized genomic regions across a sample or between samples and controls [110]. sWGS has a limit of detection of between 5%-10% mutant allele fraction [115], and so has limited sensitivity for profiling earlier stage disease. If molecular profiling of a small number of recurrent copy number alterations is desired, higher sensitivity may be achieved through targeted sequencing of single nucleotide polymorphisms, which may detect copy number alterations as low as 0.5% [116].

The limit of detection for assays will vary based on whether the individual's disease status, and tumour mutations, are already characterised. Tumour burden in plasma has often been assessed by quantifying mutations (or other alterations) that were previously identified in the patient's tumour sample [27, 28]. Prior knowledge of the mutation profile from tumour sequencing data enables the detection of known patient-specific mutations

above the background error rate, as opposed to calling mutations *de novo* [29, 111]. For mutation calling across a panel of genes or hotspots, the risk of false positives increases with the size of the panel due to multiple hypothesis testing [117], and filters need to be applied to increase specificity, which erodes sensitivity for rare variants. Thus, sequencing-based assays can be used as sensitive and quantitative tools for ctDNA measurement and monitoring, in addition to their use for mutation profiling [29, 30, 111].

ctDNA can be quantified using different metrics, such as mutant allele concentration (i.e. copies per ml) or mutant allele fraction [118]. Each of these metrics would be affected in a different way by analytical, pre-analytical, and physiological characteristics. For example, metabolic changes to the rate of ctDNA turnover would affect the concentration of mutant alleles more than the mutant allele fraction, whereas pre-analytical factors affecting release of germline DNA from blood cells would reduce the mutant allele fraction to a greater extent. Analysis of ctDNA (both fraction and concentration) [101], as well as total cfDNA and cfDNA fragmentation [3, 101], could therefore provide complementary information, and may have advantages in different applications or in combination.

1.7 Clinical utility of liquid biopsies in oncology

There is a clear clinical need for novel diagnostic and molecular tools in oncology. For example, conventional sampling methods such as needle biopsies are subject to procedural complications in up to one in six biopsies [119], difficulty in obtaining sufficient material of adequate quality for genomic profiling (reported failure rates range from <10% to >30% of cases) [120, 121], and sampling biases arising from genetic heterogeneity [109, 122–125]. Detection and monitoring of disease over time relies on body fluid-based markers that often lack specificity [126], and imaging which exposes patients to ionising radiation [127] and has limited resolution (in both time and space). Recent advances in ctDNA research highlight the potential applications of liquid biopsies at each stage of patient management

(Figure 1.3a). These potential applications primarily arise from two types of information obtainable through ctDNA analysis: quantification of disease burden, and genomic analysis of cancer (Figure 1.3b). These may be combined and/or leveraged through serial sampling in order to monitor disease burden and clonal evolution.

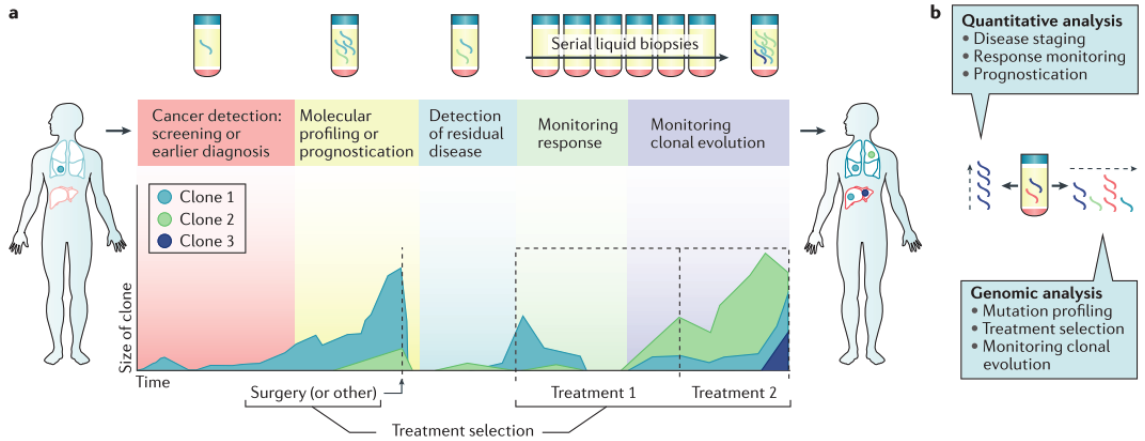


Fig. 1.3 **Applications of ctDNA analysis during the course of disease management**
(a) A schematic time course for a hypothetical patient who undergoes surgery (or other initial treatment) has a disease relapse, and then undergoes systemic therapy. The potential applications of liquid biopsies during this patient's care are indicated. The patient starts with one single disease focus, but clonally distinct metastases emerge following treatment, depicted in different colours.
(b) The information extracted from ctDNA may be classified, broadly, into quantitative information (i.e. tumour burden) or genomic information. Quantification of ctDNA at a single time point may allow disease staging and prognostication, and genomic analysis can inform selection of targeted therapies. Therefore, longitudinal analysis allows the quantitative tracking of tumour burden, such as response monitoring; and by comparing genomic profiles over time, clonal evolution may be monitored. Taken from Wan et al. [59].

1.8 ctDNA detection

The concentration of ctDNA in plasma has been shown to correlate with tumour size [62, 128] and stage [129]. A study of 640 patients with various cancer types and stages [129] found a 100-fold increase in median ctDNA concentration between patients with Stage I and Stage IV disease. Measuring individual tumour mutations in each patient, patients with Stage I disease

had fewer than 10 copies per 5 ml of plasma. In sharp contrast, patients with advanced prostate, ovarian and colorectal cancers had a median concentration of 100-1,000 copies per 5 ml of plasma. ctDNA levels vary greatly even within patients with the same type and stage of disease. This variability in ctDNA concentration is partially explained by differences in extent of metastatic spread or disease burden. In a recent report that compared ctDNA levels with tumour volume assessed by imaging in patients with relapsed high-grade serous ovarian cancer, ctDNA levels and disease volume were significantly correlated (Pearson $r = 0.59$, $p < 0.001$) [118]. Mutant alleles in plasma increased in fraction by approximately 0.08%, and in concentration by 6 mutant copies/ml, for every cm^3 of disease [118]. Despite these correlations, substantial variation in ctDNA concentration may arise from inter-individual differences. For example, poor tumour vascularisation could hamper ctDNA release into the bloodstream, or conversely, could promote ctDNA release via producing hypoxia and cell death. Histological differences could forseeably influence both the rate and type of cell death. Patients with primary brain tumours have very low levels of ctDNA, with a median concentration for individual mutations of less than 10 copies per 5 ml of plasma [129], while the fraction of tumour DNA in CSF was found to be significantly higher [45, 46, 129]. Although not directly proven, the blood-brain barrier has been suggested to impede the movement of cfDNA fragments into the circulation [45, 46, 129].

The relationship between ctDNA levels and cancer stage suggests prognostic utility for ctDNA. Patients with detectable ctDNA have been shown to have worse survival outcomes than those without [24, 130–133]. For example, the 2-year overall survival rate for patients with colorectal cancer who had detectable ctDNA was 48%, as opposed to 100% for patients without [24]. In patients with detectable ctDNA, it has been found to be a more significant prognostic predictor than commonly used tumour markers [30, 118], where an increasing concentration of ctDNA correlates with poorer clinical and radiological outcomes [30, 118, 129, 134, 135]. For example, in patients with metastatic breast cancer, a significant

inverse correlation was shown between ctDNA concentration and overall survival up to 2000 copies/ml, with a uniformly poor prognosis above this level [30]. In addition to ctDNA levels, mutational patterns identified in ctDNA can help group patients into molecular subtypes with different prognosis [136].

1.8.1 Molecular profiling of ctDNA

At present, the European Medicines Agency (EMA) and the US Food and Drug Administration (FDA) approve the use of information from ctDNA analysis to help select patients with EGFR-mutant NSCLC for gefitinib (EMA) [137], erlotinib (FDA) [90] or osimertinib [90, 138] therapy in the event that a tumour sample is not evaluable, as it provides an alternative source of material. This could offer a pragmatic solution to provide molecular profiling information for patients, while avoiding repeat biopsies for some individuals. Current recommendations [90, 138] state that if liquid biopsies are carried out in advance of a tumour biopsy, ctDNA detection may abrogate the need for tissue biopsy, but if ctDNA analysis is negative, a tissue biopsy may still provide valuable genomic information.

Molecular profiling using ctDNA may have particular utility for stratifying patients in 'basket trials', which enrol patients independent of tumour histology, or 'umbrella trials', which assign patients to multiple investigational drugs or treatment options [139]. For example, a 54-gene panel detected ctDNA in 58% of patients across multiple cancer types [107]. Of the patients with alterations, 71.4% had at least one mutation actionable by an FDA-approved drug. This panel is being used to test the feasibility of matching patients with different metastatic cancer types to targeted therapies in a prospective clinical trial. In another study presented at the 2016 Molecular Analysis for Personalised Therapy meeting, a 34-gene panel identified mutations in 79% of 174 patients with NSCLC, allowing 28 patients (17%) to receive personalised treatment based on ctDNA molecular profiling [140]. Personalised therapy selection presents challenges: even if mutations are successfully detected using

ctDNA in patients, an efficacious molecularly targeted agent may not exist. However, data from a prospective clinical trial presented at the 2016 Molecular Analysis for Personalised Therapy meeting demonstrated that selecting therapies based on genomic analysis can improve outcomes for cancer patients, even when patients with well-established actionable targets (for which approved drugs are available) were excluded [141].

Improvements in the analytical sensitivity of molecular profiling tools could further increase detection and concordance rates or allow for sensitive multiplexed analysis (Table 1.1), though biological factors and heterogeneity may reduce sensitivity in some cancer types and stages [129, 142, 143]. The utility of ctDNA should, therefore, be assessed for different clinical indications. However, benchmarking ctDNA against individual tumour biopsies may be confounded by sampling error, as rare private mutations may be sampled in the biopsy, but release insufficient ctDNA into the bloodstream to be detectable.

1.8.2 Analysis of tumour heterogeneity

The extent of genetic heterogeneity has been confirmed over recent years as multi-regional sequencing studies have demonstrated clear differences in mutation profiles between different tumour regions in the same patient [144, 145] and between different specimens from primary and metastatic sites [146]. Although the potentially confounding effects of heterogeneity are recognised, it is often neither feasible nor desirable to perform multiple tumour biopsies on patients to try to account for this. Analysis of an individual biopsy might not accurately reflect the genomic architecture of a patient's cancer, introducing bias to the selection and efficacy of personalised medicines. Furthermore, in a recent study of patients with lung cancer treated with an *EGFR* inhibitor, the tumour $EGFR^{\mathrm{T790M}}$ allele fraction correlated with the degree of tumour shrinkage [147], suggesting that the current paradigm of treatment selection based on mutation presence or absence alone may be suboptimal.

Liquid biopsies sample ctDNA released from multiple tumour regions, and may thereby reflect both intratumour heterogeneity [46, 123, 124] and spatially separated disease foci [123, 125, 148]. While individual tumour biopsies from different tumour regions may differ in mutation profile due to intratumour heterogeneity [149, 150], ctDNA analysis has detected mutations missed in corresponding tissue samples [123, 125, 142, 151]. Multi-region tumour sequencing data show that stem mutations (shared by all tumour regions) show a higher allele fraction in plasma compared to private mutations [149, 150]. Therefore, for tracking tumour burden in plasma, stem mutations would provide the most reliable detection. Alternatively, tracking a large set of mutations may compensate for potential biases of individual private mutations.

1.8.3 Sensitivity of hot-spot mutation assays and gene panels

By comparing mutation detection in plasma against matched tumour samples, the sensitivity of ctDNA analysis has been estimated in retrospective studies as between 65%-98% [92, 94, 101, 142, 148]. For profiling specific loci, for example in order to stratify patients for matched molecular therapies, international studies have begun to demonstrate that large-scale testing is feasible and may be standardised, although the use of assays with limited analytical sensitivity resulted in low detection rates of ctDNA [152, 139, 153]. Using assays developed specifically to detect low levels of ctDNA [3], a blinded prospective study demonstrated sensitivity for *KRAS* and *BRAF* mutations in metastatic colorectal cancer of 92% and 100%, respectively, with concordance rates of 96% and 100% for each [101].

Significant attention has been devoted to analysis of *EGFR* mutations in patients with NSCLC, as it is often challenging to obtain tissue biopsies to help inform treatment [119, 120]. Meta-analysis of 27 studies published between 2007-2015 [154], comprising altogether nearly 4,000 patients, resulted in a pooled sensitivity of 60% and specificity of 94% for detection of *EGFR* mutations in plasma or serum, with a variety of methods. In a phase IV study of

the *EGFR* inhibitor gefitinib, mutation status was compared between tumour and plasma samples from 652 patients. The sensitivity and specificity for detecting mutations in plasma were determined as 65.7% and 99.8%, respectively [153, 155], likely affected by the limited analytical sensitivity of the allele-specific PCR method used [89], as the version of the kit used in that study had a limit of detection (at >95% analytical sensitivity) of 1.64% and 1.26% for *EGFR* deletions and L858R mutations [153]. Low rates of concordance of $EGFR^{T790M}$ status were also observed in a recent phase III trial of osimertinib [156], in which tissue testing was compared to an allele-specific PCR assay for ctDNA analysis that has a limit of detection (with >95% analytical sensitivity) of 100 copies of $EGFR^{T790M}$ per ml of plasma [90]. Using methods based on droplet digital PCR (ddPCR), higher sensitivity in plasma was obtained. Sensitivity in plasma for the $EGFR^{T790M}$ mutation (which confers resistance to gefitinib and erlotinib, and frequently emerges following initial treatment with those *EGFR* tyrosine kinase inhibitors) was lower than the sensitivity for *EGFR* L858R and *EGFR* Ex19del, which occur earlier in the development of disease (70% vs. >90%) [142, 157]. Initial data suggests that this may be due to heterogeneous presence of resistance mutations at disease relapse [142].

In retrospective analyses, despite limited concordance rates of *EGFR* mutation status between plasma and tumour samples observed in some studies, response rates for patients who were plasma positive for mutations in *EGFR* were similar to response rates of patients who were tissue positive [142, 153, 155–157]. Data showing the response of patients treated solely on the basis of ctDNA analysis are starting to emerge: patients who were treated with osimertinib based on detection of $EGFR^{T790M}$ in plasma showed response rates similar to the those of patients treated based on tissue analysis [140]. Interestingly, objective responses were also seen in patients with very low allele fractions of mutant $EGFR^{T790M}$ in plasma (<0.5%) [140].

1.8.4 Detection of structural variants

Copy number alterations can be detected in cfDNA using WGS [31, 32, 110, 158], amplicon-based [104, 114, 159, 160], and hybrid-capture approaches [34, 111, 161]. In patients with hepatocellular carcinoma, WGS was able to identify amplifications and deletions in plasma matching those identified in tumour tissue [32, 162]. Heterogeneous copy number changes were also identified in a patient with synchronous breast and ovarian cancers, as copy number changes unique to each cancer were detected in plasma [32]. In a study of 80 patients with prostate cancer, androgen receptor (AR) copy number gain prior to abiraterone therapy predicted a worse overall survival, thus identifying patients with primary resistance [163]. For patients with advanced disease, sWGS may provide a relatively cost-effective measure of ctDNA level that is applicable across cancer types. This approach may have utility as a sample screening step in a ctDNA analysis workflow [114], where high-burden patient samples are triaged for exome sequencing [34].

Chromosomal rearrangements in plasma can be identified through both WGS [31] and hybrid-capture sequencing approaches [33, 164], though the latter may be more economical due to the depth of coverage needed to confidently identify a rearrangement. In one study of patients with prostate cancer, sWGS was able to detect a deletion on chromosome 21 in 5 patients, though higher-depth capture sequencing was necessary to identify a rearrangement between exon 1 of *TMPRSS2* and exon 3 of *ERG* [33].

1.8.5 Early detection

Diagnosing cancer at an earlier stage, particularly before metastatic spread, may allow earlier intervention and could improve survival [165]. A number of studies have demonstrated the potential for non-invasive early diagnosis. Mutations have been detected in saliva and plasma from individuals up to two years prior to cancer diagnosis [18, 166], and there have been reports of incidental pre-symptomatic detection of cancers in pregnant women who

underwent NIPT [51, 167, 168], as WGS can identify copy number alterations of both fetal and tumour origin. Screening in asymptomatic populations introduces risks of over-diagnosis and false positives; implementation could therefore be explored in stages, and a first step could involve the use of ctDNA for earlier diagnosis of disease in symptomatic individuals, who at present may undergo lengthy investigation procedures.

In a survey across several cancer types, ctDNA was detected in 82% of patients with Stage IV disease, which fell to 47% for patients with Stage I disease [129]. The method they applied was benchmarked as being able to detect one copy of an individual cancer mutation per 5ml of plasma [129]. Using a sequencing gene panel targeting a median of four mutations per patient, ctDNA was detected in 50% of patients with stage I NSCLC [111]. Targeting known tumour mutations in plasma using ddPCR assays in early-stage breast cancer showed a sensitivity of 93.3% [169]. An sWGS method adapted from an NIPT assay was recently shown to detect 6/16 (37.5%) cases of early ovarian cancer [170], though this approach may not perform as well in other cancer types with fewer copy number alterations. Together, these studies outline the possibility, and the challenge, of detection of ctDNA in early stage disease.

1.9 Disease monitoring

1.9.1 Monitoring response

The short half-life of cfDNA [28, 54, 55, 171], as well as the ease and reduced risk of repeating liquid biopsies relative to imaging [127] or tissue biopsies [119], enables liquid biopsies to be used for real-time monitoring of cancer burden in response to therapy. Studies monitoring patients during treatment have shown that ctDNA dynamics correlate with treatment response [28–30, 111, 133], and may identify response earlier than clinical detection [30, 172, 173]. In breast cancer, ctDNA showed the greatest range in concentration and

provided the earliest measure of response to chemotherapies, as well as the earliest indication of impending relapse compared to imaging and other blood-based cancer markers, such as CTCs and Cancer Antigen 15-3 (CA 15-3, also known as MUC1) [30]. In relapsed ovarian cancer, pre-treatment ctDNA levels and the extent of ctDNA decrease after chemotherapy initiation were significantly associated with time to progression, and were more informative than levels of CA125 (also known as MUC16) [118].

A recent study suggested that an early spike in ctDNA levels (allele fractions of *BRAF* mutations) in the first week following the initiation of immunotherapy in melanoma patients may predict response [174]. This may reflect a transient increase in cell death. If these data are confirmed, sampling at early time points could be applied in the clinic as well as in drug development. However, the presence or timing of such spikes in cell death would likely vary based on the pharmacological properties and biological responses to treatments used. An early spike was not observed a few days after initiation of treatment with chemotherapy for patients with colorectal cancer [175] or with an *EGFR* inhibitor for patients with NSCLC [173]. If analysis of plasma immediately after the start of therapy could reliably detect the destruction of sensitive cancer cells, this raises an exciting possibility that the existence of resistant sub-clones could be identified very rapidly through differential early dynamics of mutations. In the context of immunotherapy, liquid biopsies may provide the opportunity to monitor both ctDNA and the response of the immune system, for example through the analysis of cfDNA released from distinct T-cell clones [176].

1.9.2 Clonal evolution and resistance

As discussed above, rising or falling ctDNA concentration may provide an indication of treatment effect on overall tumour burden. If multiple tumour mutations are interrogated, then the relative change between each may provide insight into molecular evolution of the patient's cancer [34, 177, 178]. Ratios between the levels of different mutations in plasma

can indicate heterogeneity and may be informative to predict patient response to treatment targeting particular alterations [142, 161]. Liquid biopsies have been shown to contain ctDNA from multiple tumour sites [46, 123–125], can have a faster turnaround time than tissue biopsies [33, 94, 164], and may be less prone to biases resulting from individual tumour biopsies [144, 145, 179]. Studies demonstrate that ctDNA can monitor clonal evolution and identify resistance mechanisms to treatment [108, 151, 157, 161, 180]. Serial ctDNA analysis in patients with colorectal cancer demonstrates the positive selection of mutant *KRAS* clones during *EGFR* blockade, which later decline upon the withdrawal of anti-*EGFR* therapy [151, 181]. In NSCLC patients undergoing treatment with *EGFR* inhibitors, resistance-conferring mutations emerged in plasma ahead of clinical progression. Exome sequencing of plasma DNA may identify resistance mechanisms in patients across cancer types [34], though the sensitivity of exome sequencing currently limits its application to advanced cancer patients where ctDNA levels are high (>5% mutant allele fraction). Design of patient-specific mutation panels [29, 125, 129, 182] could be a more cost-effective alternative for high-sensitivity monitoring, though may miss subsequent *de novo* events. Serial sWGS analysis also demonstrates highly dynamic copy number adaptations in response to selection pressures, with a mean interval of 26.4 weeks between new amplifications [183].

Serial liquid biopsies may have particular utility for adaptive clinical management, whereby resistance mutations are prospectively identified, and therapy adapted in real-time (Figure 1.4). In the clinical research setting, non-invasive monitoring could facilitate adaptive clinical trials, which prospectively identify efficacious treatment regimens or drug combinations, and identify resistance mechanisms to novel therapies. In addition, *in vitro* or *in vivo* experiments carried out in parallel may provide greater insight into cancer biology. For example, colorectal cancer cell line experiments carried out in parallel with ctDNA analysis showed that resistance mutations may arise from both the selection of pre-existing minor clones, and through ongoing mutagenesis [177]. Another study investigating resistance

878 to a pan-tropomyosin-related kinase (TRK, also known as NTRK) inhibitor in colorectal
879 cancer demonstrated that simultaneous analysis of patient-derived xenografts and liquid
880 biopsies may characterise resistance more comprehensively than plasma alone [184].

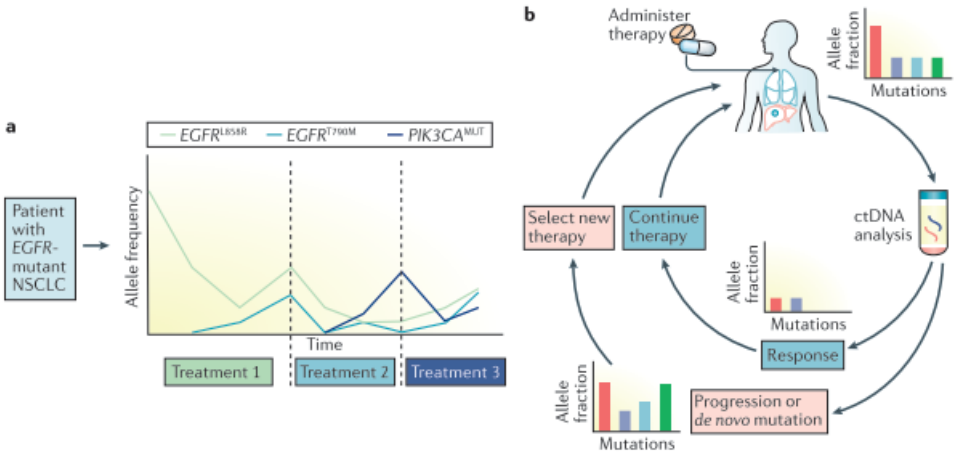


Fig. 1.4 **Adaptive or reactive treatment paradigms using liquid biopsies**
(a) During systemic anti-cancer therapy, serial liquid biopsies may identify biochemical response or progression. If progression is identified, the clinician may be able to switch therapy, or select a therapy to target arising clones carrying additional mutations that were identified by this analysis.
(b) This adaptive or reactive monitoring and treatment may continue as a loop, which would be facilitated by a fast turnaround time for ctDNA analyses, for example through the use of point-of-care diagnostics. The timeframes for this analysis can vary between hours and months; the former could allow analysis of early kinetics in response to therapy. Taken from Wan et al. [59].

881 1.9.3 Minimal residual disease detection

882 Following surgery or treatment with curative intent, even in the absence of any other clinical
883 evidence of disease, detection of ctDNA may signal the presence of minimal residual disease
884 (MRD), which could identify patients who may be at a higher risk of relapse. Stratification of
885 patients into high- and low-risk groups would enable adjuvant therapy to be given to patients
886 who stand to benefit most, while sparing low-risk patients from unnecessary morbidity
887 and risks of adverse events. In a prospective study of 230 early-stage colorectal cancer

patients [185], assessment of ctDNA within 4-10 weeks after surgical resection indicated that recurrence-free survival at 3 years was 0% for the ctDNA-positive and 90% for the ctDNA-negative groups. In a separate study of 55 patients with early-stage breast cancer [186], assessment of ctDNA showed that detection of ctDNA at first follow-up could also indicate poor prognosis in early-stage breast cancer. Furthermore, stratification based on mutation detection across serial samples improved prediction of relapse, and this and other studies have observed an interval of 7.9-11 months between ctDNA detection and clinical relapse [186, 187], similar to that identified in the metastatic setting [30].

Instead of SNVs, patient-specific DNA rearrangements identified from sequencing tumour samples can be used to design assays to track tumour burden in plasma [184]. Curative surgery could provide an excellent opportunity to obtain tumour DNA that can be sequenced to guide the design of assays for post-operative monitoring. Patient-specific rearrangements may be detected in ctDNA at levels as low as 0.001% mutant allele fraction [188, 189], since rearrangements are less confounded by background noise. One of the challenges of individualised panel design is that sequencing an individual tumour biopsy may not sample every mutation in heterogeneous disease; therefore, sequencing matched body fluid and tumour samples may be desirable for comprehensive mutation profiling. In future, if tumour sequencing were to become more routine, monitoring disease using patient-specific panels might become viable, although regulatory issues of standardisation and assay validation around the design and delivery of such assays may be complex.

With increasingly sensitive approaches, earlier and more accurate identification of patients likely to relapse might become possible, thereby facilitating treatment decision making. If levels of ctDNA were to rise sufficiently, molecular profiling for *de novo* mutations may be carried out, in order to identify molecular targets for intervention.

1.10 Enhancing detection of residual disease

1.10.1 Targeting multiple patient-specific mutations

Following the initiation of chemotherapy or following surgery, ctDNA concentration may decline to undetectable levels using single-locus assays, often due to there simply being no mutant molecules (at the locus of interest) in the plasma sample collected. Even with a perfectly sensitive assay, the probability of detection of ctDNA decreases as ctDNA concentration declines, as any single mutation of interest may not be present in a given volume of sample. At low ctDNA concentrations, due to sampling error, some mutations will be detected while others are missed (Figure 1.5). The number of DNA molecules interrogated by an assay thus places a hard limit on the maximum achievable sensitivity. An increase in the number of molecules interrogated must be accompanied with a reduction in background error rates in order to detect signals occurring at a lower frequency; this may be achieved through bioinformatic methods (Figure 1.6a).

It should be noted that for cancers with a viral aetiology, e.g. nasopharyngeal carcinoma or cervical cancer, detection of the cancer-associated viral DNA that may be present in body fluids in many more copies than tumour DNA can enhance the identification of individuals with early stage disease or pre-malignant lesions with a high risk for cancer [32, 190, 191]. Chan et al. [191] recently demonstrated in a cohort of 20,174 individuals that screening for Epstein-Barr Virus cfDNA can identify a group of patients that are at significantly increased risk for nasopharyngeal carcinoma (70% vs. 97% PFS at 3 years; HR 0.1).

Sampling multiple pre-specified mutations in each reaction may improve detection of low levels of ctDNA, since every additional target provides another independent opportunity to test for the presence of a mutant molecule from the set of DNA molecules at that locus [29, 106, 113]. Prior knowledge of the mutation profile from tumour sequencing data enables the detection of known patient-specific mutations, and reduces the number of false positive

mutation calls from mutation calling across a panel of genes. By doing so, fewer filters have to be applied in order to improve specificity of such calls, thereby maximising sensitivity. Taken to a (currently impractical) extreme, ultra-deep sequencing of the entire genome could in the future allow sensitive detection of cancer even from small volumes of plasma [31]. However, targeting many mutations can only improve the limit of detection to the level of the background error rate. Therefore, in order to achieve the highest sensitivity, it is important to both maximise the number of molecules available, and reduce background error rates from PCR and sequencing.

1.10.2 Maximising recovery of molecules

To increase the number of molecules available for analysis, the simplest solution may be to collect larger volumes of blood, though the number of blood tubes acceptable to take at once would eventually become limiting. This may potentially be overcome by analying multiple body fluids in parallel (Figure 1.6b), such as urine for bladder cancer [42, 192] or stool for colorectal cancers [16, 17]; or cytological specimens such as cervical smears [193], uterine lavage [194] , or oesophageal brushings [195] for gynaecological or oesophageal cancers, respectively. Alternatively, methods such as plasmapheresis or implanted devices containing materials that bind cfDNA might provide a solution, and similar devices have been tested for enhancing the yield of circulating tumour cells (CTCs) [196]. Alternatively, if it is acceptable and practical for the patient to have multiple blood samples taken over multiple days, this could further boost the number of cfDNA molecules available for detection of rare mutant molecules.

Given the low number of cfDNA molecules present in an individual plasma sample, analytical methods should aim to minimise cfDNA molecule loss. Molecules may be lost at every stage of library preparation, particularly during the ligation of adaptors onto starting molecules or library cleanup steps [197]. To maximise library preparation efficiency,

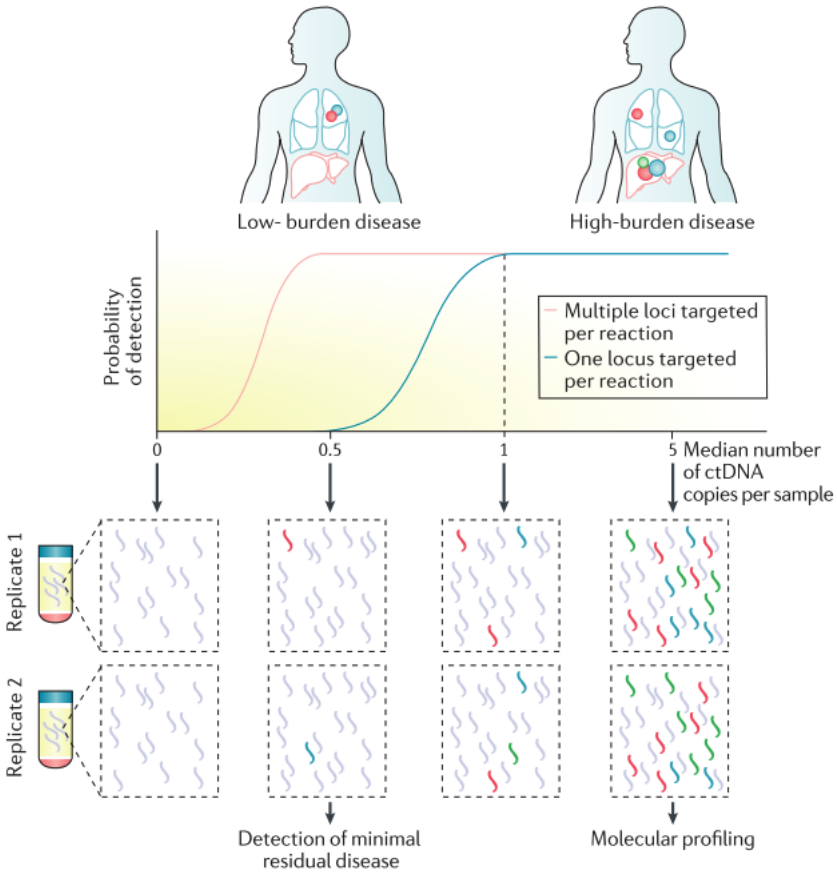


Fig. 1.5 **Leveraging multiple mutations to detect low-burden disease and overcome sampling noise**
Even with a perfectly sensitive assay, the probability of detection of ctDNA decreases as ctDNA concentration declines, as any single mutation of interest may not be present in a given volume of sample. At low ctDNA concentrations, due to sampling error, some mutations will be detected while others are missed. Sampling multiple pre-specified mutations in each reaction may improve detection of low levels of ctDNA, since every target provides an independent opportunity to test for the presence of a mutant molecule in the set of DNA molecules at that locus [29, 106]. Boxes below the graph show hypothetical examples of sets of molecules that may be captured by each replicate in the analysis of a sample. Taken from Wan et al. [59].

Newman et al. [111] increased adaptor ligation time from 15 minutes to 16 hours with a 100-fold molar excess of adaptors. As a result, recovery efficiency was improved by >300% to recover 49% of starting molecules, calculated based on the ratio between the number of unique molecules identified during sequencing compared to the number of input molecules, calculated based on qPCR [111].

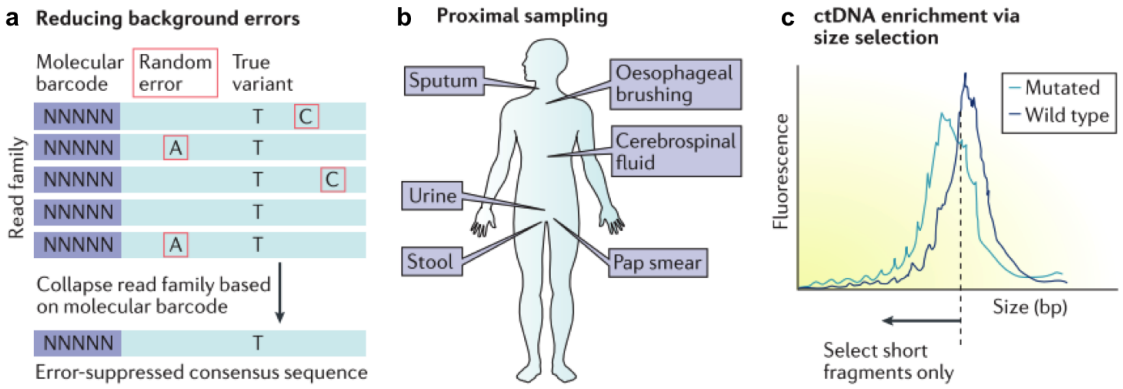


Fig. 1.6 **Methods to improve detection of ctDNA**
(a) For molecular barcoding, unique molecular sequences are added to each molecule during library preparation so that sequencing reads originating from the same starting molecule can be identified. By comparing all reads from the same molecule, a single consensus sequence can be taken, which can suppress errors arising from PCR or sequencing.
(b) To improve sensitivity of analysis, for example for disease diagnosis or detection of MRD, other body fluids may be considered in combination with, or instead of, plasma. Sampling of body fluids or cytological specimens proximal to the tumour site may yield a higher concentration of DNA of tumour origin.
(c) ctDNA has been shown to be shorter than cfDNA [64, 65, 68, 198]. Thus, selection of shorter fragments experimentally or in silico may enrich for sequences of cancer origin [65, 199] and can improve sensitivity for samples with low fractions of ctDNA. Taken from Wan et al. [59].

1.10.3 Error-suppression

Background error rates

Background noise in sequencing data arises from PCR and sequencing errors, hampering detection of low levels of ctDNA even when there are sufficient molecules present to overcome

971 sampling error. During PCR, errors arise from two main sources: DNA-polymerase-catalysed
972 enzymatic copying, plus errors due to DNA thermal damage (e.g. A+G depurination, oxida-
973 tive damage, and deamination effects) [200]. Extension errors arising from DNA polymerases
974 can be mitigated through the use of polymerases that have 3' editing activity, though they
975 are limited by their slower extension rate: *Pyrococcus furiosus* has an extension rate of 20
976 nucleotides per second, whereas *Thermus aquaticus* elongates at 80 nucleotides per second.
977 Using molecular barcoding [112] to suppress sequencing errors, error rates of PCR have
978 been estimated at 4.5 x 10^{-7} errors/bp/PCR cycle [201].

979 During Illumina sequencing, library molecules serve as the template for a PCR reaction
980 whereby a complementary sequence is synthesised using nucleotides bound to fluorophores.
981 By labelling each type of base with its own fluorophore, the synthesised sequence can be
982 identified using imaging; by having multiple templates imaged in parallel, the term Massively
983 Parallel Sequencing arises. Errors can arise from the synthesis of library molecules becoming
984 out of phase, for example due to the lack of a terminator on a newly added base [202]. In
985 addition, another error can occur if termination is not reversed during the washing cycle.
986 Overall, any error due to early or late base termination during synthesis can cause a phasing
987 error. To account for errors in Illumina sequencing, for every base sequenced, it is given an
988 associated base quality score which quantifies the probability of it being an error.

989 Illumina sequencing has been benchmarked in a large metagenomic dataset as having
990 SNV error profiles of 2 x 10^{-3} and 4 x 10^{-3} in reads 1 (R1) and 2 (R2), respectively [201].
991 Indels occur at a much lower frequency compared to SNVs (2.8 x 10^{-6} and 5.1 x 10^{-6}, for
992 R1 and R2, respectively), in part due to multiple bases having to be incorrectly read by the
993 sequencer to produce an indel call. SNV error rates vary both by the type of mutation (i.e.
994 mutation class) [111, 106, 201], position in the read [201], and sequence context [201]. Base
995 qualities of substitutions were characterised by Schirmer et al. [201], which showed that

996 69-86% of SNVs showed low base quality scores, indicating that they were likely sequence
997 artefacts.

998 To mitigate the effects of sequencing error, stringent base quality filters have been applied
999 for the early detection of disease across multiple cancer types using ctDNA without prior
1000 mutation knowledge [203]. It is possible that multiple rounds of read alignment (using
1001 different alignment algorithms) may reduce mapping errors, as demonstrated by Chan et al.
1002 [204], who used the Burrows–Wheeler transform and Smith–Waterman-like algorithms in
1003 parallel, and compared the mutations from each. Furthermore, reads can be quality-trimmed
1004 to remove lower quality bases at the end of the read, particularly R2 [201, 205]. Regarding
1005 sequence context, particular trinucleotide base contexts appear to be more common than
1006 others, with 'GGG' and 'CGG' being the first and second most noisy motif in each dataset
1007 analysed by Schirmer et al. [201].

1008 **Methods of error-suppression**

1009 PCR and sequencing background noise can be suppressed molecular barcoding (Figure 1.6a),
1010 enabling ctDNA to be detected at allele fractions below 0.1% [106, 112, 206, 104]. Molecular
1011 barcoding uniquely identifies each starting molecule during the ligation step in library
1012 preparation, so that each original molecule can be reconstructed using sequencing data.
1013 During the reconstruction of each original starting molecule, error-suppression is achieved
1014 by taking the consensus read sequence across multiple redundant sequencing reads of the
1015 same molecule (the read family), which has a unique molecular barcode and a unique set of
1016 genomic co-ordinates (start and stop locations).

1017 Additionally, during the ligation step of library preparation, if starting molecules are
1018 ligated to adaptors which identify each of the + and - strands of the original double-stranded
1019 DNA molecule, upon sequencing the original duplex can be reconstructed [99]. In addition
1020 to generating a consensus across the read family members, a consensus can be generated

between the + and - strands, thereby reducing the error rate to a theoretical level of 3.8 x 10^{-10}. For first-round PCR errors to be incorporated into the consensus sequence, the mutation would need to occur at the same base position on both strands of the dsDNA duplex, *and* result in complementary errors. This approach has been applied to ctDNA analysis [106], though the low error rate is difficult to fully utilise given the challenging of sequencing sufficient molecules: starting with 5ng input (1650 genome equivalents) and a conversion rate of 50%, if only one locus is targeted, having a limit of detection of better than 1 x 10^{-3} would not provide any additional sensitivity. In order to increase the number of molecules interrogated further, larger input or a larger number of mutations should be targeted, with greater sequencing performed.

Instead of adding molecular barcodes, dilution or partitioning of molecules enables the identification of unique molecules. For example, Hoang et al. [207] devised the *bottleneck sequencing system* whereby molecules are diluted to such an extent that start and end positions of each molecule (endogenous barcodes) can be used to uniquely identify each molecule. By ligating forked Illumina adaptors to each molecule, each dsDNA duplex can be recreated based on endogenous barcodes only. This approach enabled genome-wide analysis of mutation rates in normal human tissue, but may not be utilised to interrogate specific loci due to the limiting dilution of molecules which resulted in low or no coverage for any given locus. Rosenfeld et al. [206] suggested an alternative approach of splitting each cfDNA sample to be analysed into multiple replicates, each with a small number of molecules (e.g. 40 molecules per reaction). By segregating molecules through partitioning, any mutant molecules present would be observed at a high allele fraction in that partition, whereas noise would occur at lower mutant allele fractions (except for first-cycle PCR errors). Thus, this partitioning approach is analogous to molecular barcoding, in that each partition and the observed AF provides information on the original starting molecules. The challenge of this approach is of physically achieving a large number of replicates, since if too few

replicates are performed then the total number of molecules interrogated is low and again limits sensitivity.

1.10.4 Size selection

Since mutant molecules are shorter than wild-type cfDNA molecules [198], the size of a cfDNA fragment may be leveraged to enhance detection of mutant molecules (Figure 1.6c). Given that mutant molecules show a modal length of ~147bp [65], such molecules can be selected for either through enrichment of short molecules *in vitro*, *in silico* [199], or by attributing short molecules greater weight during analysis.

In the NIPT field, size selection has been applied to enrich for fetal cfDNA [208], which increased the fetal DNA fraction, but simultaneously reduced the number of total DNA molecules being counted, hampering sensitivity for confident detection of aneuploidy. Later, Chan et al. [204] utilised a dynamic cut-off whereby fetal fragments were only considered for genotyping of fetal-specific SNPs if the mean fragment size of mutant (candidate fetal) molecules was at least 10bp shorter than the wild-type (maternal) fragments. Taken together, it seems that for genotyping individual loci, any approaches that take into account fragment size should take caution to not cause excessive allelic loss. This may potentially be achieved through upweighting the signal from short fragments at each locus, or through size *exclusion* approaches instead of size selection, or by interrogating a large number of loci through patient-specific sequencing approaches.

Size selection has been shown to produce enrichment and enhanced detection of ctDNA in patients with advanced ovarian cancer using sWGS [199]. sWGS is ideally suited to size selection due to the molecules of interest not being limited to one particular locus, and so any loss of mutant molecules is more acceptable. This combination of techniques pushes the limit of detection of sWGS potentially up to an order of magnitude [199]. While this

approach is broadly applicable, it may still only detect ctDNA in patients with relatively high levels of disease burden.

1.10.5 Biological limits on sensitivity

Although the above technical advances may improve sensitivity for ctDNA analysis further, biological and genomic factors will eventually become limiting. For confident *de novo* cancer detection using ctDNA, detected alterations should have a high positive predictive value for cancer. However, mutations known to be associated with cancer (e.g. in *TP53*, *KRAS*, and Notch pathway genes) have been found at low levels in skin biopsies of healthy individuals [209]. If non-tumourigenic clones were to increase to sufficient size and release mutated cfDNA, they could introduce biological noise. Clonal haematopoiesis with leukaemia-associated mutations has been observed in 10% of individuals older than 65 years of age, though the absolute risk of conversion to haematologic cancer is 1% [210]. Genomic alterations known to be associated with cancer have been found in plasma from healthy individuals [106, 107, 210]. Clinical outcomes for apparently healthy individuals in whom mutant DNA is detected in plasma should be characterised in order to understand the biological and clinical implications of such findings.

1.11 Future directions for the field of liquid biopsy

Proof-of-concept studies provide an excellent starting point for larger prospective studies into the clinical utility of ctDNA, and demonstrate that ctDNA may be a useful research tool for drug development, and for the study of intratumour heterogeneity and clonal evolution. Moving forwards, randomised trials comparing ctDNA-guided decision-making against the standard of care would be definitive, and the EMA have outlined good practice guidelines for the design of such trials [211]. Trials to test the clinical utility of ctDNA analysis for

treatment monitoring are now being carried out [212]. In one trial, patients with NSCLC receiving erlotinib are being prospectively monitored, and if resistance mutations emerge in plasma, then additional scans to search for signs of disease progression would be carried out [212]. Another clinical trial aims to demonstrate the efficacy of targeting mutations identified in plasma from patients with advanced breast cancer [213], which could support the future use of plasma-only mutation profiling and treatment stratification. Together, these studies highlight that the field is moving from exploratory ctDNA studies, towards clinical trials where ctDNA is guiding decision-making.

A better understanding of the origin and biology of cfDNA and ctDNA would aid the implementation of liquid biopsies [35]. The relative contributions of apoptosis, necrosis and active release, particularly at different time points during treatment, should be explored. Our limited understanding of the release and clearance mechanisms of cfDNA hampers interpretation of current studies. Studies of the dynamics and reproducibility of ctDNA measurement in the absence of intervention will become increasingly important as we aim to interpret ctDNA signal in response to treatment. It is also not clear whether all tumour subclones contribute proportionately to the total ctDNA pool, or whether their representation in the bloodstream is biased by other biological factors, such as tumour vascularity or metabolic activity. *In vivo* cellular barcoding experiments [214] and autopsy studies [125] could elucidate the contribution of individual subclones, and histological studies may clarify the factors that modulate ctDNA release. The differences in size between cfDNA and ctDNA fragments [51, 64, 65, 198] suggest that optimising processing and extraction methods (as well as downstream assays) for recovery of short fragments may provide further improvement to overall performance.

While ctDNA can have greater sensitivity and specificity compared to other circulating biomarkers [30], taking a multi-marker approach may offer a more comprehensive insight into a patient's disease [101, 198, 215]. For example, total cfDNA concentration correlates

with disease status [2, 198] and is associated with prognosis [216]. Epigenetic analysis of cfDNA may identify cancer gene hypermethylation [24, 32] or the cell-type giving rise to cfDNA fragments [49–51], and may provide a window into the tumour microenvironment, which usually lacks somatic mutations. Other circulating nucleic acids such as mRNA and microRNA can provide additional layers of information [217]. Targeting multiple types of nucleic acid, with independent mechanisms of release, may increase sensitivity for detection of MRD, for example through the co-isolation of both exosomal RNA and cfDNA [218]. Actively released nucleic acids may be preferred for the detection of mutations in subclones resistant to therapy, whereas fragments arising from dying cells following the initiation of therapy may identify treatment-responsive subclones. Next, although it may be possible to infer gene expression patterns from cfDNA [183], sequencing RNA within exosomes [219], CTCs [220] or platelets [221] could provide more direct evidence. We further echo the suggestion by Gormally, Hainaut and colleagues [222], made almost a decade ago, that characterisation of proteins associated with cfDNA may provide a rich source of information on an individual's disease, and the biology of cfDNA.

The clinical uptake of liquid biopsies will depend on the practical advantages for patients and clinicians, the infrastructure required, and its cost-effectiveness. Tissue biopsies will continue to play a key role in cancer management, particularly for the histological diagnosis and classification of cancers. At present, specialised laboratories handle CTC and ctDNA samples [223], though in future hospital laboratories may carry out analysis locally if appropriate processes can be established [224].

Point-of-care devices for the identification of individual hotspot mutations with clinically meaningful sensitivities are starting to be used for tissue and plasma samples [225, 226]. The feasibility of single molecule (third-generation) sequencing of maternal plasma DNA was first demonstrated in 2015 [69], and subsequently it was shown that structural variants in cell line DNA can be detected [227]. The portability of such technologies was demonstrated by

the real-time genomic surveillance in the field during the Ebola virus disease epidemic [228]. At present, such platforms are limited by a high error rate [227], making single nucleotide variant and indel detection challenging. Another challenge is that of sequencing short DNA fragments, which requires optimised library preparation methods [229]. Furthermore, sequencing capacity is also limited (~150 megabases) [230], though this is likely to increase in the near future, and specific amplicons may be targeted through real-time selective sequencing [231]. These studies support the possibility of molecular profiling at the point of care, especially if blood plasma can be interrogated without the relatively cumbersome and time-consuming step of DNA purification [6].

The initial approvals by the EMA and FDA for mutation detection in plasma as a companion diagnostic [90], and emerging ctDNA-guided clinical trials [212, 213], represent key milestones towards the implementation of liquid biopsies in personalised oncology. Improving technologies are enabling an ever-wider scope for non-invasive molecular analysis of cancer, providing information that opens new avenues for genomic research and may aid in clinical decisions. In order to fully exploit the potential utility of liquid biopsies, it is essential that the biology of ctDNA be explored further. Thus far, liquid biopsies have demonstrated the potential for utility across a range of applications, and are beginning to be used for patient benefit.

Chapter 2

Current progress in monitoring melanoma with liquid biopsies

2.1 Abstract

Circulating tumour DNA (ctDNA) is being increasingly used to detect and monitor disease in patients with both early-stage and advanced cancers. In cutaneous melanoma, liquid biopsies are beginning to be used in research settings to monitor response to targeted therapies and immunotherapies, and to monitor for relapse following surgery. A number of studies have demonstrated that ctDNA can monitor advanced melanoma, though sensitivity is limited for early-stage and low tumour-burden disease. By virtue of having a high mutation rate in melanoma, additional mutations can be targeted, which may offer enhanced sensitivity. Recent developments, and the barriers to clinical implementation of ctDNA, will be discussed in this review.

2.2 Epidemiology of melanoma

Cutaneous melanoma is a type of skin cancer affecting melanocytes (pigment producing cells). In 2014, there were 15,400 new cases of melanoma in the United Kingdom, and its incidence is rising faster than any other cancer in the UK [232]. Cutaneous melanoma causes 75% of deaths related to skin cancer, and its global incidence is 15-25 per 100,000 individuals per year [233].

Increasing sun (UV) exposure is an important contributor to this rise in incidence, supported by the high rate of UV damage-induced C>T transitions in the genome of melanomas [234]. As a result, melanomas have the highest mutation rate per megabase of all cancers, with an average of 16.8 mutations/Mb [234].

91% of individuals diagnosed with melanoma are identified at an early stage i.e. American Joint Committee on Cancer (AJCC) Stage I or II, which is characterised by the presence of localised disease without regional lymph node involvement or distant metastasis (Table 2.1). Patients' outcomes are influenced by Breslow thickness, ulceration and regional nodal and metastatic disease status. Stage I-II melanomas are managed primarily by surgical excision, which confers 5-year survival of >98%; this falls to 62% for patients with Stage III disease, i.e. disease that has spread to nearby lymph nodes (Table 2.2). For patients diagnosed with Stage IV disease, 5-year survival has historically been approximately 10% [235], though recent evidence using combination immunotherapy indicates that three-year overall survival rates can be as high as 60% [236].

2.3 Management

2.3.1 Stage I-III melanoma

Surgical removal is the gold standard treatment for patients with primary cutaneous melanoma with negative regional lymph nodes [239]. Following surgical resection of localised disease,

Stage	Description
Stage 0	Melanoma is confined to the epidermis.
Stage IA	Melanoma thickness up to 1mm AND no ulceration AND it has a mitotic rate of less than 1 mitosis per mm^2
Stage IB	Melanoma thickness of up to 1mm with ulceration, OR it has a mitotic rate of greater than 1 mitosis per mm^2, or it is between 1-2mm and is not ulcerated
Stage II	Melanoma thickness greater than 1mm with ulceration, or greater than 2.0 mm without ulceration.
Stage III	Regional lymph node involvement and/or presence of microsatellites, satellites or in-transit metastases, with no evidence of distant metastases
Stage IV	Distant metastatic spread of melanoma.

Table 2.1 **Melanoma Staging Classification**. Adapted from the AJCC 8th Edition [237].

Stage	Men	Women
I	101.9%	100.4%
II	78.2%	85.3%
III	50%	54.7%
IV	8.4%	25.2%
Stage not known	74.2%	79.9%
All stages	85.8%	92.2%

Table 2.2 **Melanoma 5-year relative survival statistics**. Taken from the Office of National Statistics 2016 [238].

patients undergo surveillance for tumour recurrence and additional primary melanomas (which may occur in up to 8% of patients [240]).

During surveillance of patients with stage IA disease, any skin lesions that are new or have changed, or that violates any of the ABCDE (Asymmetry of the lesion, Border irregularity, Colour variability, Diameter >6 mm, Evolving) rule, should be evaluated by a dermatologist or specialist, and/or removed and sent for histological assessment. For disease stage IB-III, history and physical examination should be carried out every 3-6 months for 3 years, then 4-12 months for 2 years, then yearly [241]. The survival benefit from early detection of melanoma recurrence was demonstrated in a study by Leiter et al. [242], which showed a significantly improved survival among patients whose recurrent metastases were detected early (40.5% overall survival) rather than late (25.6% OS, P=0.013).

Adjuvant immunotherapy may be provided for patients with stage III or definitively treated stage IV disease and can produce significant benefit, though they carry risks of side effects and negatively affect quality of life. In 2008, Bottomley et al. carried out a study on a cohort with stage III melanoma after full lymphadenectomy, where patients were randomised to either observation (n = 629) or pegylated interferon alfa-2b for an intended duration of 5 years [243]. After 3.8 years of median follow-up, recurrence-free survival (RFS) risk was reduced by 18% (HR = 0.82) compared to observation alone, though at a cost of significantly lower global health-related quality of life (-11.6 points) based on the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire C30) [243].

Between 2008-2011, 951 patients with stage III cutaneous melanoma were randomly assigned to ipilimumab, an anti-CTLA4 immunotherapy, or placebo [244]. At a median follow-up of 2.74 years, the median RFS was shown to be 26.1 months in the ipilimumab group versus 17.1 months in the placebo group (HR 0.75). Although patients gained significant benefit from the treatment, adverse events led to the discontinuation of treatment of 245 (52%) of the 471 patients who started ipilimumab, and 5 patients (1%) died due to

drug-related adverse effects. Most recently, Weber et al. [245] compared adjuvant therapy with nivolumab versus ipilimumab single-agent treatment in patients undergoing resection of stage III or IV melanoma. In this study of 906 patients, Weber et al. [245] demonstrated that adjuvant nivolumab resulted in a significantly longer RFS and a lower rate of grade 3 or 4 adverse events than ipilimumab.

2.3.2 Stage IV melanoma

Prior to the Food and Drug Administration (FDA) approval of ipilimumab in 2011, the main treatments for stage IV melanoma were chemotherapeutic agents such as dacarbazine and temozolomide [239], despite single or combination chemotherapies having never been shown to improve overall survival. The introduction of targeted therapies led to improvements in overall survival over the past decade [246]. Most recently, immunotherapies have shown potential to produce long-term response [239]. Thus, immune checkpoint inhibitors and targeted therapies have now largely superseded chemotherapy treatment for cutaneous melanoma, although both show high rates of toxicity [239].

Anti-*BRAF* inhibitors such as vemurafenib and dabrafenib are widely approved for treatment of patients with *BRAF* $V600^{mut}$ metastatic melanoma [247]. In randomised clinical trials, Vemurafenib (960mg twice daily) [246] and dabrafenib (150mg twice daily) [248] both showed similar response rates and improvements in PFS, reducing the risk of progression by >60% relative to dacarbazine chemotherapy. In the vemurafenib trial, the risk of death was reduced by 63%. The median overall suvival with dabrafenib was 20.0 months vs. 15.6 months for dacarbazine alone, and 13.6 months vs. 9.7 months for vemurafenib compared to dacarbazine, respectively. With targeted therapies, common adverse effects observed were skin-related toxicity, arthralgia and fatigue [239]. In 20% of patients treated with *BRAF*-inhibitors, secondary cutaneous squamous cell carcinomas and keratocanthomas can

occur within the first 2-3 months of therapy [249]. In addition, second primary melanomas occur in 2-5% of patients treated with anti-*BRAF* drugs [239].

The efficacy of ipilimumab has been demonstrated by randomised clinical trials of ipilimumab compared to placebo [250], and as compared to a gp100 peptide vaccine, which targets a glycoprotein that is highly expressed in melanocytic cells [251]; these studies showed that compared to gp100, ipilimumab showed a hazard ratio of 0.66. However, one of the challenges of treatment with immunotherapy is the frequency and severity of adverse events. In a study of 676 patients with unresectable stage III or IV melanoma, grade 3-4 immune related adverse events occurred in 10-15% of patients with colitis, skin rash and endocrinopathies being common [251]. When ipilimumab is combined with dacarbazine, the rate of grade 3-4 adverse events rises 56% [239]. For patients who show serious adverse reactions, it is recommended to discontinue ipilimumab and to treat with high-dose corticosteroids.

Monitoring of immunotherapy response by imaging is made particularly challenging by unique response patterns of such treatments, termed pseudoprogression [252]. It was noted in early immunotherapy cases that patients who showed conventional disease progression based on RECIST (i.e. an increase in the sum of measures of target lesions, unequivocal increase in non-target disease, or the appearance of new lesions) were found to have late but durable treatment responses [253]. As a result, systematic criteria for assessing immune-related responses were defined beyond those described by RECIST or WHO criteria [254], which were named as Immune-related Response Evaluation Criteria In Solid Tumors (irRECIST).

For patients with metastatic melanoma, disease monitoring must be able to overcome the challenges posed by the now widespread use of immunotherapy, such as mitigating the confounding effects of pseudoprogression, and having sufficient dynamic range to quantify both high levels and residual disease levels if long-term response were to be achieved.

2.4 Treatment monitoring with ctDNA

2.4.1 Stage IV melanoma

Liquid biopsies for ctDNA have been taken serially to monitor advanced melanoma alongside conventional markers [133, 134, 172]. The aim of using liquid biopsies for melanoma is for earlier identification of treatment response or failure, which may enable modification of clinical management. ctDNA may have advantages over conventional serum biomarkers due to its wide dynamic range [30] and high specificity [59]. Currently, patients are monitored with imaging and prognostication is also aided by serum protein markers such as lactate dehydrogenase or S100B [255]. S100B proteins, which are shed by melanoma cells, have value in monitoring patients during therapy, with increasing levels correlating with progressing disease (and vice versa) [256], and have a higher specificity for progressive disease than LDH [257]. While protein markers can be informative for disease monitoring, they are less sensitive than ctDNA analysis for disease progression or for non-RECIST progression events [258], and do not reflect individual mutations in the tumour.

Aside from protein markers, ctDNA has been shown to correlate with imaging data, as measured by FDG-PET ($r = 0.61$; $P<0.001$) [259, 260]. Patients were monitored using NGS and dPCR for mutations in *BRAF*, *NRAS*, *KIT*, or *TERT*. Targeting *KIT* or *TERT* facilitates monitoring of *BRAF*, *NRAS* and *NF1* wild-type melanoma patients [261], which can contain these additional hotspot mutations. McEvoy et al. [260] found that ctDNA was non-detected in all patients with a low level of tumour lesion glycolysis across all lesions as measured by PET/CT, highlighting that greater sensitivity that what was achieved in this study is needed to detect and monitor small lesions or tumours with low levels of metabolic activity. Despite this limited sensitivity, an assay with limited sensitivity can still have utility as a prognostic classifier, as McEvoy et al. found that non-detected ctDNA is associated with longer PFS

[260]. These studies indicate the potential for liquid biopsies to provide a minimally invasive monitoring option between and in addition to imaging.

Immunotherapy

Liquid biopsies have been used to monitor patients with immunotherapy [133, 134, 262]. In both immunotherapy and chemotherapy patients, based on digital PCR measurement of *BRAF* or *NRAS* mutations in a cohort of 48 patients with metastatic melanoma, lower pre-treatment ctDNA levels were significantly associated with response to treatment and prolonged PFS [133]. The dynamics of ctDNA in response to therapy appear to differ based on treatment type: targeted therapy caused a steep decline in ctDNA [172], which was not the case with immunotherapy [133]. These differences in ctDNA dynamics are likely related to the mechanism of action of each of the treatments and the timeframes of cancer cell death. Following treatment initiation with dabrafenib and trametinib, patients whose *BRAF* $V600^{mut}$ ctDNA remained detectable showed significantly poorer PFS than patients whose ctDNA became non-detected [172]. Rising ctDNA was able to identify clinical progression on targeted therapy either at (26% of patients) or before (44%) progression [172].

By virtue of the high specificity and short half-life of ctDNA [59], liquid biopsies may be able to identify pseudoprogression in patients receiving immunotherapies, which can be observed as a clinical progression but followed by ctDNA levels immediately declining [134]. A recent study comparing irRC and ctDNA in 125 patients showed that ctDNA profile may predict pseudoprogression from true progression with a sensitivity of 90%, and 100% specificity [262]. Compared to LDH, ctDNA showed a positive predictive value of 100% vs. 92%. Thus, ctDNA analysis may have utility for classification of patients with pseudoprogression.

2.4.2 Stage I-III melanoma

Following radical treatment of stage I-III melanoma, recommended follow up is every 3-6 months for up to 5 years with history, examination and imaging [263]. More frequent surveillance imaging is performed for patients with stage III melanoma [263]. Stratifying patients using ctDNA may enable more appropriate selection of patients for adjuvant treatment, and could complement AJCC staging information. This is particularly relevant given the evidence to support the use of ipilimumab, nivolumab and dabrafenib/trametinib for Stage III melanoma [264, 245, 265].

In the early-stage disease setting, Lee et al. [266] used ddPCR to assess ctDNA in 161 patients with high-risk Stage II/III melanoma patients following surgery. Using a threshold of $\geq$1 mutant ctDNA copies per sample from up to 2ml of plasma, 15/132 (11%) and 4/29 (14%) of patients had detected levels of *BRAF* and *NRAS* ctDNA within 3 months of surgery. Patients with detected ctDNA had a significantly reduced disease-free interval (Hazard Ratio, HR 3.12, P<0.0001), and showed significantly worse overall survival at 5 years (HR 2.63, P=0.003). Although detection of residual disease is possible following surgery in melanoma, detection rates are low, due to limited technical sensitivity relative to the biological levels of ctDNA.

Despite the low detection rate of ctDNA immediately post-surgery, assays with limited sensitivity can still classify samples into those with the highest ctDNA levels vs. non-detected ctDNA and thus can still have potential for prognostic utility. Studies with longitudinal follow-up [185, 186] show that sensitivity for patients who relapsed is higher where multiple longitudinal samples are used to classify samples. By using longitudinal samples, such studies leverage a larger number of molecules and allow time to elapse, enabling ctDNA levels to rise until they cross the limit of detection of that assay. Thus, only by increasing the sensitivity of the assay at a given time point used can the lead-time to relapse be improved.

This may partially explain why studies have shown lead-times up to 11 months but no further [117, 186, 187, 267].

2.5 Improving detection of relapse

Although single-locus digital PCR assays have low false-positive rates, in order to fully utilise the sensitivity of the assay, a sufficiently large number of input cfDNA molecules needs to be used. Frequently used Bio-Rad ddPCR systems are capable of generating up to 20,000 droplets, with an average of 70% droplets read (i.e. 30% dead volume) [268]. In theory, one mutation molecule per 20,000 can be detected depending on the background noise of the assay. Thus, to achieve sensitivity to 1 mutant molecule in 14,000, approximately 67ng of cfDNA is required as input, which would require the entire material from two blood tubes, given a median cfDNA concentration of 7.8ng/ml plasma in the plasma of melanoma patients [269].

Sampling error may be overcome, to an extent, by targeting additional mutations using targeted sequencing. However, fixed gene panels may still achieve limited sensitivity in the plasma of melanoma patients, as there are few recurrently mutated loci in the melanoma genome [121]. For example, Siroy et al. [270] applied a 46-gene sequencing panel to tumour tissue samples from 699 melanoma patients, identifying a median of 1 mutation per patient, with >1 mutation being observed in 1/3 of patients [270]. If such a panel were to be applied to plasma, many patients' ctDNA level cannot be quantified due to having few or no informative reads overlying loci that were mutated in that patients' tumour (i.e. the issue is of sampling error and not analytical sensitivity). To boost the median number of mutations detected in plasma, even larger sequencing panels would be required. However, regardless of the panel size used, the background error rate of the assay will ultimately set the limit of sensitivity. By designing patient-specific sequencing panels [29, 106, 113], greater sequencing depth can be achieved at loci of interest.

While there has been rapid progress in monitoring of melanoma using liquid biopsy, there is a need for greater sensitivity in order to accurately classify patients into high- and low-risk of relapse following surgery. In addition, non-invasive cancer diagnostics may potentially improve the classification of patients for adjuvant therapy, which may reduce the rate of overtreatment with immunotherapies that carry risk of adverse events. This information may guide treatment decisions by clinicians and patients, in order to maximise patient outcomes and quality of life.

2.6 Conclusion

There is increasing evidence supporting the validity of using of ctDNA for monitoring both early and advanced-stage melanoma. The majority of these studies use single-locus mutation assays targeting *BRAF* or *NRAS*. Despite the limited sensitivity of single-locus assays, ctDNA has still been shown to be able to prognosticate in both high- and low-burden disease settings. As liquid biopsies become more sensitive, potentially through targeting additional mutations present in the melanoma genome, disease may be detected earlier. In order to progress these findings towards implementation, future clinical trials guided by liquid biopsies are needed to test whether or not the real-time, sensitive and specific nature of ctDNA may produce clinical benefit for patients.

Chapter 3

Targeting multiple mutations in plasma improves genotyping and monitoring

3.1 Attribution

This chapter is adapted from the following manuscript, which we submitted in November 2016. The submitted manuscript from 2016 is as follows:

Wan JCM, Murphy S, Gale D, Morris J, Mouliere F, Marass F, Heider K, Chandrananda D, Smith CG, Bignell GR, Alifrangis CA, Parkinson C, Durrani A, McDermott U, Massie C, Corrie PG, Rosenfeld N*.
Plasma DNA analysis mitigates biased mutation detection in tumour biopsies.
*Corresponding author

Using a patient-specific sequencing approach, the first half of the manuscript suggests the potential advantage of using a combined plasma and tumour data for comprehensive genotyping, and the latter half discusses the potential improvement in sensitivity of detection of ctDNA by targeting multiple mutations.

3.2 Author Contributions

I performed all analyses in this Chapter. The first half of this chapter on comparison of mutation profiles comprises the manuscript which we submitted in 2016, which I wrote and performed all analyses, and was given feedback from co-authors. The second half of the manuscript focuses on longitudinal monitoring of ctDNA by using multiple patient-specific mutations with TAm-Seq.

3.2.1 Methods line-by-line contribution

- **Patient cohort and sample processing**. Pippa Corrie and Nitzan Rosenfeld set up the MelResist sample collection. Individualised ctDNA monitoring was suggested in a study previously published in the lab [29], and so this study was carried out to explore individualised sequencing data using TAm-Seq to both improve sensitivity, and to study tumour heterogeneity.

- **DNA extraction**. Tumour DNA extraction was performed by Graham Bignell and Ultan McDermott at the Sanger Institute, and plasma DNA extraction and quantification was shared between myself and Suzanne Murphy.

- **Tumour exome sequencing**. Graham Bignell and Ultan McDermott performed exome sequencing of tumour samples of 10 stage IV melanoma patients for a separate hypothesis.

- **Amplicon sequencing**. Suzanne Murphy (wet-lab MD student) worked with Francesco Marass and James Morris (bioinformatics staff) to design individualised TAm-Seq (iTAm-Seq) primers for these patients based on the previously generated exome data. I worked with Suzanne Murphy to perform the experimental work for TAm-Seq on plasma samples using these primer pairs.

- **Shallow Whole Genome Sequencing**. The sWGS pipeline was set up in our lab by Dineika Chandrananda and Chris Smith based on published methods [271], which Dineika applied to whole-genome libraries I generated from these longitudinal samples.
- **Mutation detection**. I developed these criteria for detection of ctDNA at multiple loci given background error rates at each locus.
- **Calculation of mean allele fraction**. Here, I suggested the use of a depth-weighted mean allele fraction as a measure of quantification of ctDNA across multiple patient specific mutations. This concept of aggregation/averaging of signal across multiple mutations was built upon in the subsequent INVAR manuscripts.

3.3 Abstract

Circulating tumour DNA (ctDNA) is increasingly used to both genotype disease and monitor tumour responses. Single-locus assays and fixed sequencing panels generate data on individual mutations only, resulting in molecular profiling of a limited number of mutations, and with a limited overall sensitivity due to sampling error for rare mutant molecules. Patient-specific approaches allow interrogation of multiple mutant loci, and also greater sensitivity for ctDNA. Here, we designed patient-specific sequencing panels targeting a median of ~50 mutations per patient, which we applied to longitudinal plasma samples from stage IV melanoma patients undergoing treatment with chemotherapy or immunotherapy. First, we show discordance in mutation profiles between matched tumour and plasma samples, indicating the potential utility of plasma mutation profiling alongside tumour profiling. Next, we demonstrate that the limit of detection of amplicon sequencing in plasma can be improved to 0.01% by aggregating signal across multiple shared and private mutations rather than considering each individually. By doing so, the agreement between ctDNA and imaging was improved, and allowed identification of progression 28 days earlier than serum lactate

dehydrogenase, a currently used protein marker. We suggest that interrogation of multiple shared and private tumour mutations in plasma can help to mitigate sampling error for tumour genotyping, and can boost sensitivity for disease monitoring.

3.4 Introduction

Shared and private mutations in heterogeneous disease, and from multiple tumour regions, can be detected in patient blood plasma samples as circulating tumour DNA (ctDNA) [125, 124]. Mutations that are shared between multiple tumour loci appear in plasma at higher mutant allele fractions, as opposed to those that are private to individual lesions [125, 124]. Serial measurement of ctDNA has previously been applied to non-invasively monitor melanoma patients [258, 134, 172, 272], though following the initiation of chemotherapy or following surgery, ctDNA concentration may decline to undetectable levels using single-locus assays, often due to there simply being no mutant molecules (at the locus of interest) in the plasma sample collected.

Patient-specific plasma sequencing approaches can enable a larger number of mutations to be interrogated in plasma. Thus far, 10-32 mutations per patient have been targeted [113, 106]. Patient-specific approaches provide greater sensitivity for ctDNA by interrogating a larger number of loci [59], and generate high-depth plasma sequencing across multiple mutations, enabling the comparison of tumour and plasma mutation profiles in plasma.

In this study, we designed patient-specific TAm-Seq panels [29], each targeting between 48-65 mutations per patient, for 9 patients with stage IV melanoma. Both shared and private tumour mutations were included in the panel design based on sequencing of multiple tumour biopsies, which enabled assessment of mutation representation in plasma at high disease burden time points, and improved detection of ctDNA overall at low disease burden time points. Based on discordance between tumour and plasma mutation profiles with high-depth sequencing, we suggest that a combined tumour biopsy and liquid biopsy approach may

provide a more accurate genotype than a tumour biopsy alone. Furthermore, we suggest that targeting multiple shared and private mutations in plasma with patient-specific panels can boost sensitivity for low-burden disease by providing additional opportunities to sample rare mutant molecules.

3.5 Results

3.5.1 Targeting multiple mutations reveals sampling biases in tumour biopsies

Mutation profiles were compared between tumour and plasma, before and after treatment, for 9 patients with metastatic melanoma (Supplementary Figures 3.10- 3.18, page 74). Across all patients, we found a Pearson correlation of 0.58 between tumour and plasma mutant allele fractions ($P < 2.2 \times 10^{-16}$), consistent with published data [124]. Data from one patient (MR1012) are shown in Fig. 3.1: for this patient, two metastatic biopsies were collected from the left groin, one collected at baseline, prior to treatment initiation with dabrafenib and trametinib, and another biopsy collected 190 days later at progression (Fig. 3.1A). Allele fractions for 39 mutations were measured in each sample in both tumour TAm-Seq and plasma TAm-Seq. Mutation loci showed a greater change in allele fraction in longitudinal tumour biopsy samples compared to longitudinal plasma samples (Fig. 3.1B). The differences between baseline and progression time point allele fractions can be quantified by assessing the coefficient of determination values (R^2) determined by linear regression (Fig. 3.1C), which showed smaller differences in allele fraction relative to the overall trend in plasma as compared to tumour. In addition, we observed that high-depth amplicon sequencing reduces the variability in tumour allele fraction as compared to exome sequencing (Fig. 3.1C), which is summarised in Supplementary Fig. 3.9. These data indicate that the mutation profiles

obtained from tumour mutation biopsies differ compared to plasma samples, which could be due to tumour biopsy sampling error.

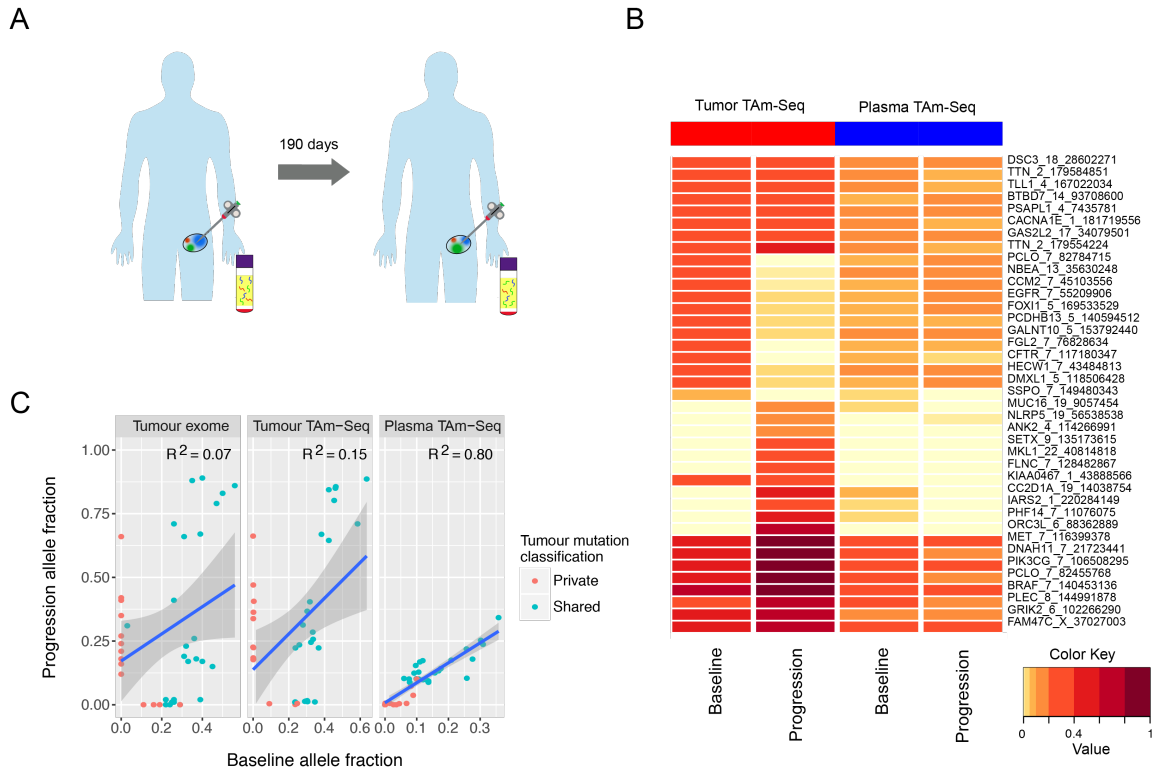


Fig. 3.1 **Comparison of tumour and plasma mutations from matched samples**
Panel A shows the sampling timeline for this patient: metastatic biopsies and plasma samples were taken, 190 days' apart, both from metastases to the left groin. This patient received dabrafenib and trametinib during the intervening period.
Panel B shows a heat map with mutant allele fractions in tumour and plasma DNA samples at baseline and progression. Mutations are ordered based on hierarchical clustering of allele fractions across tumour and plasma samples.
Panel C shows baseline mutant allele fractions plotted against progression mutant allele fractions, for tumour DNA sequenced with exome sequencing and TAm-Seq, and for plasma DNA sequenced by TAm-Seq. A linear regression line is shown, and its adjusted R^2 value is indicated.

Tumour mutations were classified as shared if they were any mutant reads in both tumour samples, otherwise the mutation was classified as private to one tumour biopsy. Patient-specific mutation detection is described in Method 3.8.1. The median allele fraction of shared mutations was significantly greater than that of private mutations when measured in either

tumour and plasma samples (P<0.001, Welch's Two Sample t-test, Fig. 3.2A), reflecting their increased intratumour prevalence. Shared mutations showed a greater difference in mutant allele fraction in plasma compared to in tumour samples, enabling greater accuracy in classification of mutations as shared or private based on mutant allele fraction alone (Area Under the Curve, of 0.89 ± 0.02 vs. 0.70 ± 0.02, P<0.001, Fig. 3.2B).

96.1% of shared tumour mutations were detected in time-matched plasma, whereas 37.5% of private tumour mutations were detected in plasma (P=0.002, Wilcox test). In addition, an average of 3% (1.2 mutations out of 40) were detected in plasma at a time point when they were not observed in the tumour sample. The total number of detected mutations are shown in Fig. 3.2C. These data were confirmed by re-analysis of data from Murtaza et al. [34], which showed that plasma mutation detection sensitivity increases as the prevalence of that mutation increases within the tumour bulk (Fig. 3.2D), and mutations identified in >2 metastatic biopsy sites were detected in plasma with a median sensitivity of 100%. To determine a plasma detection rate for mutations in the Murtaza et al. [125] study, only one tumour and plasma sample pair was used, as this was the only time-matched sample pair. Mutations with >1% mutant allele fraction (MAF) were classified as detected. Mutations were classified as shared if they were detected in >1 tumour sites.Taken together, these data suggest the utility of performing matched plasma sequencing in parallel to any tumour genotyping in order to allow identification of sampling biases for private mutations of low-intratumour prevalence, depicted in Fig. 3.3.

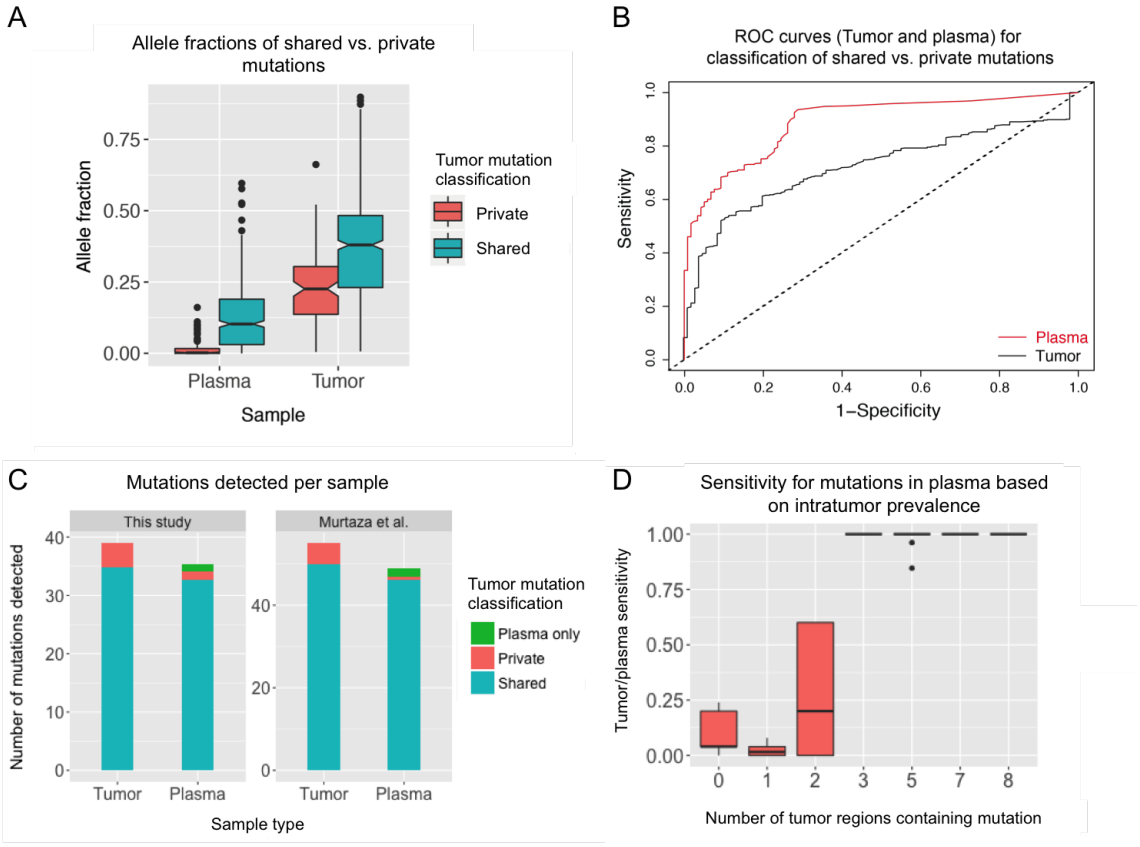


Fig. 3.2 **Plasma mutation representation**
Panel A shows box plots of allele fractions in plasma and tumour for tumour-private and tumour-shared mutations.
Panel B shows Receiver Operating Characteristic curves for the classification of shared and private mutations in tumour and plasma. The Area Under the Curve values for tumour and plasma were 0.70 ± 0.02 and 0.89 ± 0.02, respectively ($P < 0.001$).
Panel C shows the mean number of detected mutations per matched sample in the current study of Stage IV melanoma patients, as compared to the data from a case study of a patient with metastatic breast cancer. 96.1% of tumour-shared mutations were detected in time-matched plasma, whereas 37.5% of tumour-private mutations were detected. An average of 1.2 mutations (out of 40) per patient were detected in the plasma at a time-point where these were not detected in the tumour analysis.
Panel D contains box plots showing the detection rates for mutations present in a varying number of metastatic sites, using data from Murtaza et al. [125]. Mutations were classified as shared if they were present in both tumour biopsies in this study, or if they were present in at least 5 sites in the Murtaza et al. study [125].

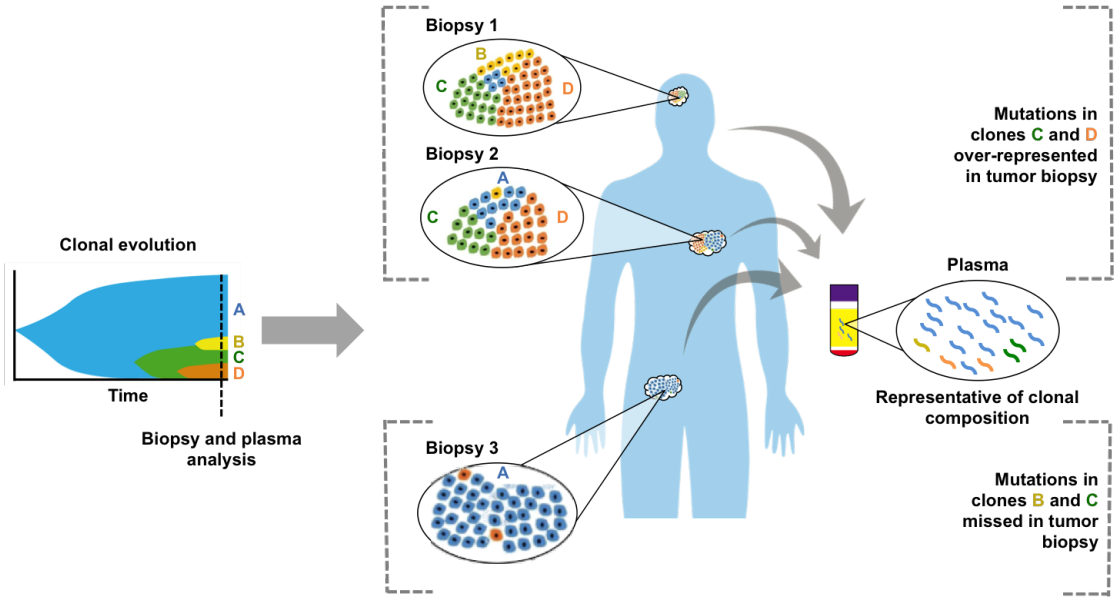


Fig. 3.3 **Plasma DNA analysis mitigates biased mutation detection in tumour biopsies** The clonal composition of this hypothetical patient's disease is shown on the left, and the time point of tumour and plasma sampling is indicated. Individual biopsies may each give a different mutation profile due to sampling error in spatially heterogeneous disease. Matched plasma analysis can reveal the extent of this sampling bias, identifying mutations that are relatively under- or over-represented in plasma. By comparing the two modalities, the effect of sampling error may be minimised.

3.5.2 Benchmarking the sensitivity of patient-specific TAm-Seq

Next, we assessed the sensitivity of patient-specific TAm-Seq for individual and aggregated mutations for disease monitoring. We applied patient-specific TAm-Seq to a spike-in dilution series, made from a patient plasma cfDNA eluate serially diluted into cfDNA from healthy individuals (Method 3.7.7). Detection of patient-specific variants above background noise was performed as described in Method 3.8.1. This patient had a *BRAF* V600E mutation, which was detected in the dilution series to an allele fraction of 10^{-2} (Fig. 3.4), but was not detected in subsequent serial dilutions.

When multiple mutations are targeted, signal can be aggregated by taking an average allele fraction across all loci, allowing quantification of disease to 10^{-4} mean allele fraction (Fig. 3.4). Although the individual detected mutations had allele fractions $>10^{-3}$, the depth-weighted average allows quantification of ctDNA despite having few mutant molecules present (Method 3.8.2). Quantification of ctDNA may be achieved to 10^{-4}, although due to sampling error, the mean allele fraction deviated from the expected allele fraction by up to an order of magnitude (Fig. 3.4). Aggregating signal across multiple loci is essential when the level of signal is below the sampling limit of any individual locus, since the signal can occur at any of the loci. 10^{-4} mutant allele fraction is approaching the limit of detection set by the background rates of non-error-suppressed sequencing that we characterised (Fig. 3.5).

3.5.3 Detection rates in longitudinal clinical samples

First, detection of ctDNA was assessed at baseline with individual driver mutations. Using either *BRAF* or *NRAS* mutations only, ctDNA was detected at baseline in 7/9 patients (Fig. 3.6). Patient information is shown in Table 3.1. Across all longitudinal time points, ctDNA was detected in 62 out of 104 time points (59.6%), and of the detected samples, *BRAF* or *NRAS* ctDNA levels ranged from 0.005 to 0.63 mutant allele fraction.

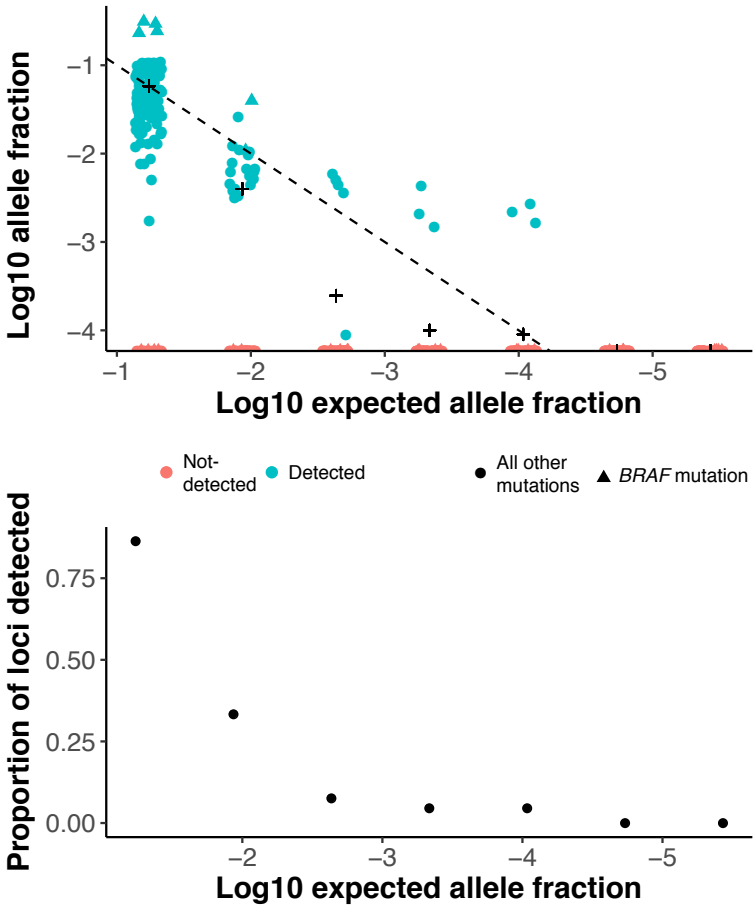


Fig. 3.4 **TAm-Seq dilution series with multiple mutations**
Patient-specific TAm-Seq was performed with 49 unique amplicon pairs, sequenced to a median depth of 66,204x. Detection of ctDNA at each locus was based on a threshold set using control samples (Method 3.8.1). The *BRAF* V600E driver mutation is indicated with a triangle, whereas all other mutations are indicated by circles. Mutations that were detected are coloured in blue, and non-detected mutations are coloured in red and plotted along the x-axis. + indicates the depth-weighted mean allele fraction across the patient-specific mutations. The dashed line indicates the expected allele fraction for each dilution level.

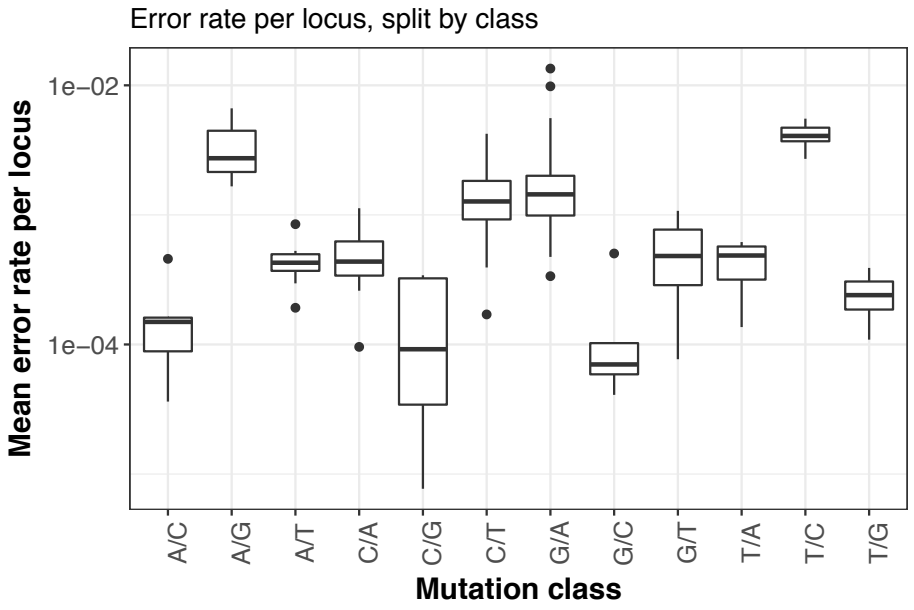


Fig. 3.5 **TAm-Seq background error rates by mutation class**
For each mutation locus interrogated, its mutation class was determined and the error rate per locus has been plotted with grouping by class. Background error rates by class varied between 0.43% to 0.0052% for T/C (T>C) and G/C (G>C) mutations, respectively.

When using multiple mutations, detection at baseline increased to 100% (9/9) patients (Fig. 3.7), and across all time points, ctDNA detection increased to 100 out of 104 (96.1%), with a range of individual allele fractions per locus of <0.01% to 33%. Although each of the mutations showed a similar overall trend, targeting multiple mutations provided a larger number of opportunities for mutant fragments to be sampled, shown by different loci showing signal and supporting detection at different time points (Fig. 3.7).

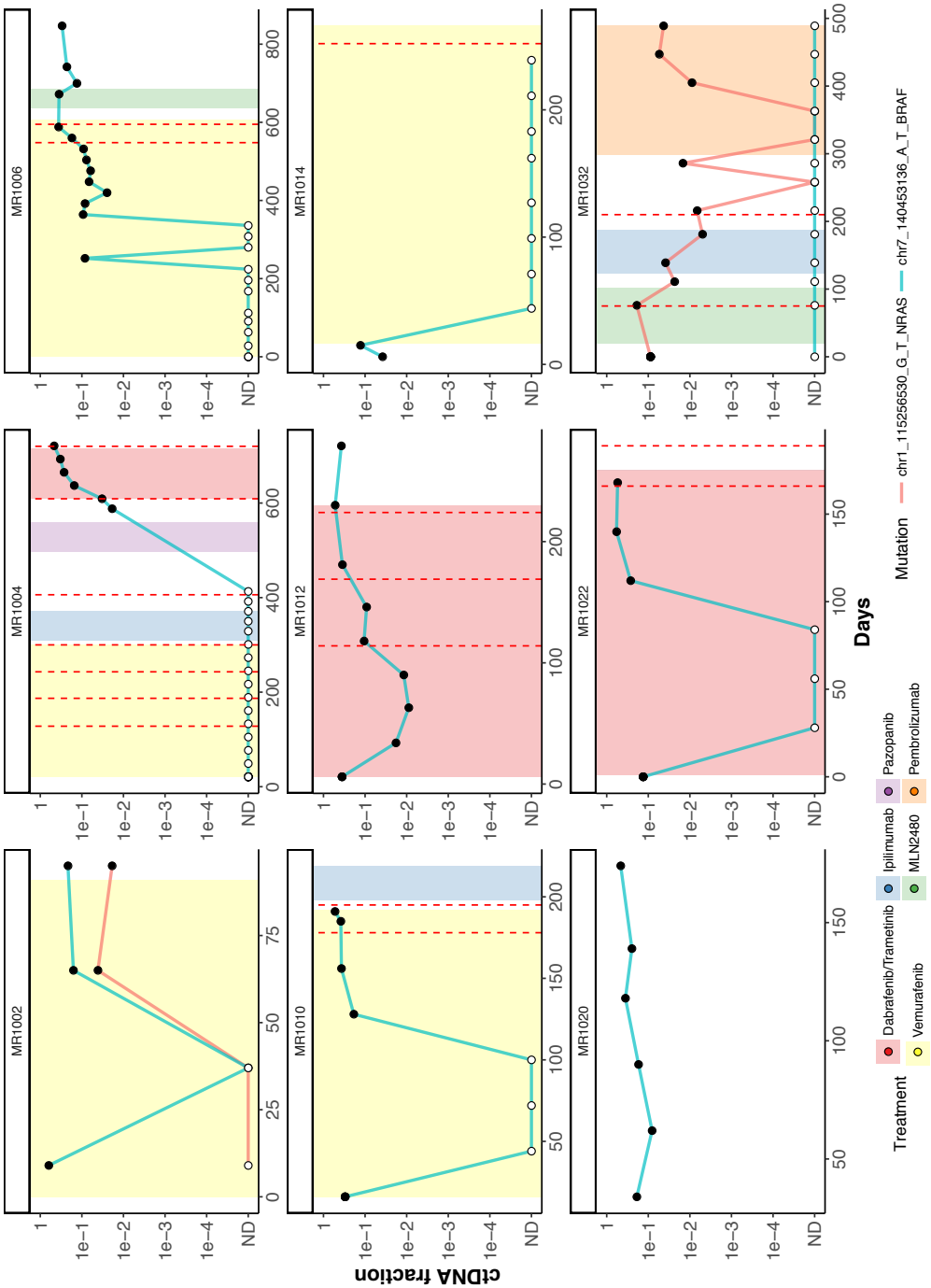


Fig. 3.6 **TAm-Seq longitudinal monitoring - driver mutations only**
For each patient with stage IV melanoma, ctDNA was assessed longitudinally at either the *BRAF* V600E or *NRAS* Q61K loci only. The time of clinical progression, as determined by RECIST criteria from CT imaging, is indicated with a vertical red dashed line.

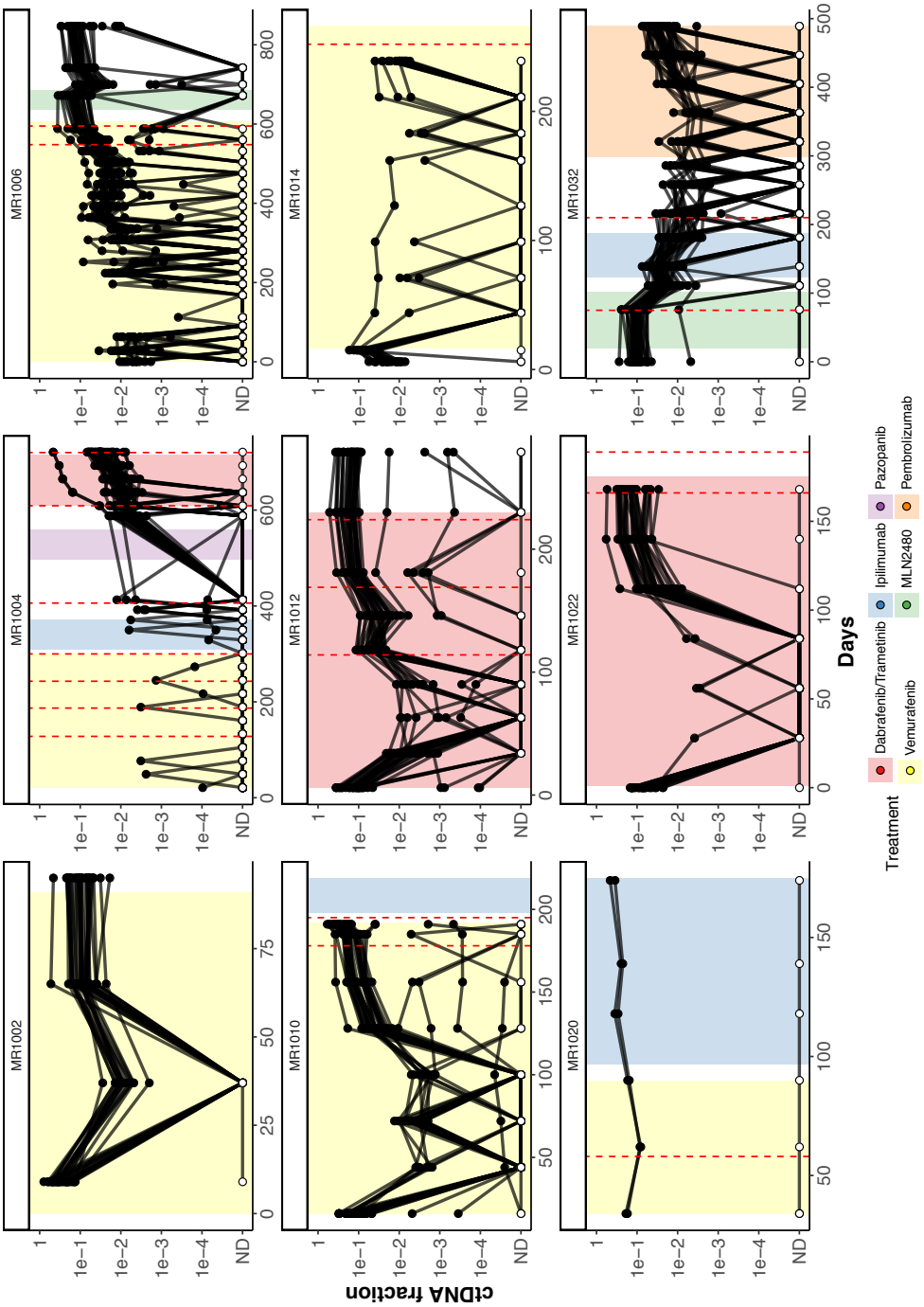


Fig. 3.7 **TAm-Seq longitudinal monitoring - all mutations**
For each patient with stage IV melanoma, ctDNA was assessed longitudinally at all patient-specific loci targeted. Detection was performed as in Methods 3.8.1. Here, a logged y-axis is used, showing the sampling of rare mutant molecules as individual loci being detected at certain time points but not others, leading to signal alternating between detection and non-detected, as in patient MR1032.

3.5.4 Comparison of ctDNA with clinical data

At time points of progressive disease, average allele fraction across all mutations was increasing immediately before at 15 out of 19 events (79%), non-detected at 2/19 events (11%), and decreasing at 2/19 events (11%, Fig. 3.8). In contrast, when one mutation was targeted, ctDNA was identified as rising in 12/19 events (63%), non-detected in 5/19 events (26%), and falling in 2/19 events (11%). For the non-detected time points in patient MR1004, there was clinical progression due to new lesions identified on CT, though the ctDNA level was below 0.01% mutant allele fraction. For the discordant time points relative to clinical data, one patient (MR1020) showed a decrease in ctDNA from 8% to 5% allele fraction at the same time as clinical progression, and patient MR1032 showed a decrease in ctDNA of 0.9% to 0.7% ctDNA at the same time as clinical progression. These data suggest that small decreases in allele fraction might still occur at progression. A larger cohort may allow better characterisation of small changes, or broader genomic sequencing may identify loci that were in fact rising in mutant allele fraction.

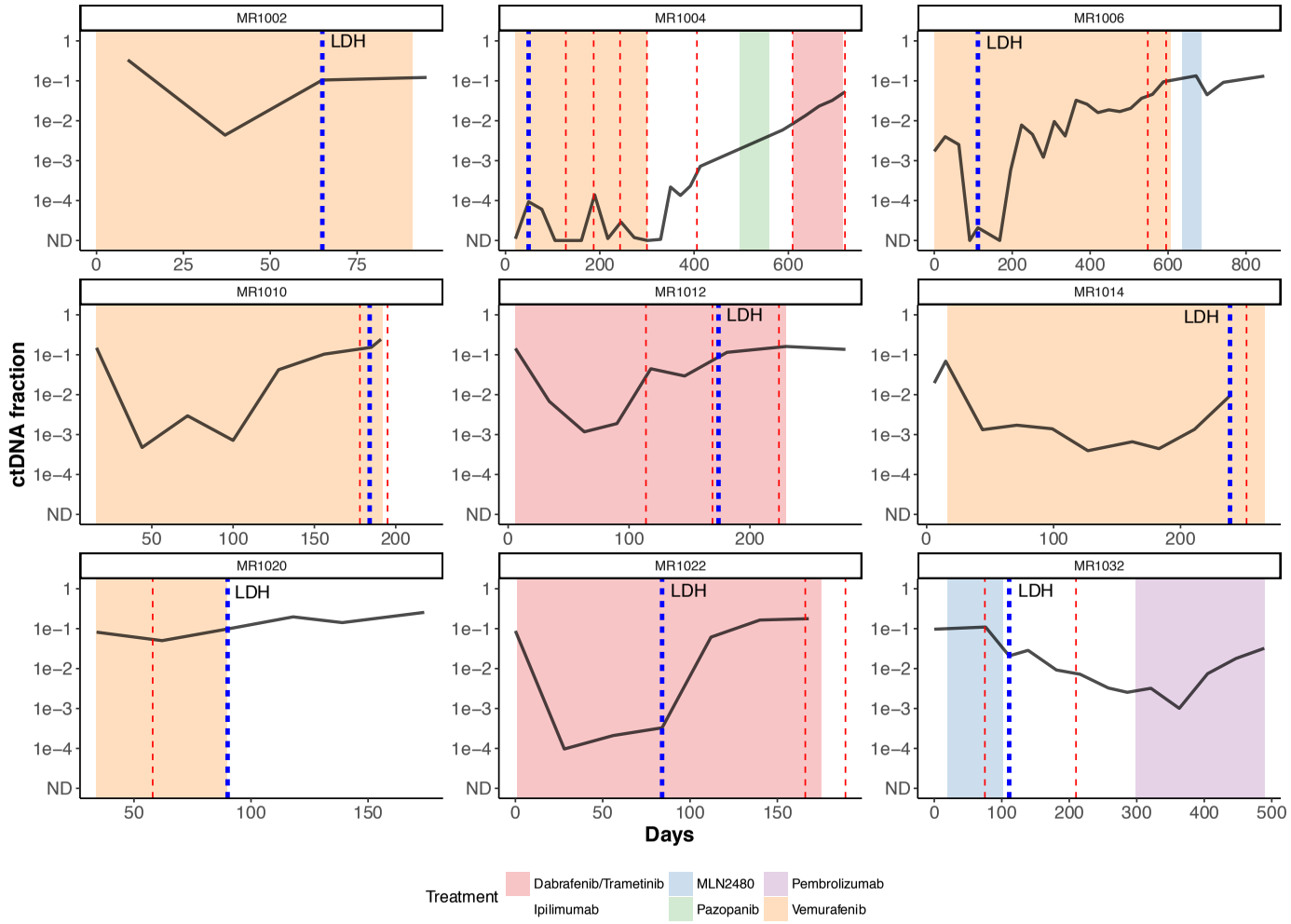


Fig. 3.8 **TAm-Seq longitudinal monitoring - average allele fraction**
An average allele fraction was calculated as the depth-weighted mean allele fraction across the patient-specific loci, as per Method 3.8.2. Treatments are indicated by shaded boxes. Time points of progressive disease are indicated by vertical red dashed lines, and the time point of LDH rising above the upper limit of normal is indicated by a vertical blue dashed line for each patient.

Following treatment with targeted therapies (either vemurafenib or dabrabrafenib/trametinib combined), ctDNA declined initially in all patients, reaching a minimum ctDNA level after a median of 62.5 days (Fig. 3.8). When ctDNA subsequently increased, driver mutations showed the highest allele fractions (Fig. 3.7), consistent with *BRAF* amplification being a known mechanism of reactivation of the *MAPK* pathway [273] (Supplementary Fig. 3.19).

Serum lactate dehydrogenase (LDH) is a melanoma tumour marker in advanced disease, with an upper limit of normal of 250IU/L. LDH concentrations were measured for each patient longitudinally. Compared to LDH, ctDNA was identified as rising a median of 28 days earlier (IQR 0-36 days, Fig. 3.8), as measured from the date of the sample where the rise was evident (with no subsequent declines in ctDNA prior to the date of the LDH rising above the limit of normal). LDH plots are shown in Supplementary Fig. 3.20.

3.6 Discussion

In this study, we show that targeting multiple shared and private mutations in plasma may have utility for both tumour genotyping and disease monitoring. We suggest that when genotyping patients with heterogeneous disease, it may be advantageous to also sequence their ctDNA at the same mutation loci in order to mitigate sampling error effects in the tumour biopsy. In addition, we demonstrate the sensitivity benefit of targeting multiple mutations in parallel to overcome sampling error for rare mutant molecules in plasma.

In our data, baseline and progression samples were temporally separated, and thus clonal evolution may influence comparisons between mutant allele fractions, such as in Fig. 3.1C. We sought to control for temporal changes by analysing multi-region sequencing studies from patients with advanced [46, 125] and localised cancers [149, 150]. We found that although tumour-based genomic profiles may vary based on the individual biopsy analysed, the biases affecting plasma sampling appear constant (Fig. 3.2). Mutations were observed in either the tumour and plasma samples that were not observed in the other sample type, at that time point. We conclude that a combined tumour and plasma genotyping approach may provide a more accurate tumour profile than either sample type alone. In this analysis, although patient-specific primer pairs had to be designed in order to verify tumour mutations in plasma, in future, it may be possible to use untargeted approaches such as plasma exome or whole-genome sequencing at high depth to generate equivalent data, if costs were to decline sufficiently.

Furthermore, targeting additional shared and private mutations in addition to the driver mutation, such as *BRAF*, provides more opportunities for rare mutant molecules to be sampled. By considering signal across multiple patient-specific loci, the limit of detection was improved to 0.01% average mutant allele fraction, approaching the theoretical limit of detection set by the background error rates of TAm-Seq (Fig.3.19). We found that a patient-specific approach allows more accurate tracking of disease relative to imaging, and

was rising 28 days prior to LDH. Although ctDNA changes either preceded or tracked progression events in most cases (79%), there were some discordant events, potentially due to RECIST scoring a progression event when a new lesion is identified while the overall burden of disease was decreasing. In 11% of events, ctDNA was non-detected, and thus the analytical sensitivity of this method was insufficient to accurately assess the change in ctDNA mutant allele fraction.

Patient-specific amplicon sequencing is being increasingly performed [113, 117]. As tumour and plasma sequencing costs decline, patient-specific plasma analysis for tumour profiling and detection of low levels of ctDNA may have increasing utility in both clinical and research settings.

3.7 Experimental Methods

3.7.1 Patient cohort

We evaluated tumour biopsy and plasma samples from nine patients with metastatic melanoma who were enrolled in the MelResist study (REC number 11/NE/0312) before and after treatment with targeted therapies and/or immunotherapy. MelResist is a translational study of response and resistance mechanisms to systemic therapies of melanoma. Table 3.1 shows clinical information for each of the patients analysed.

Patient	Baseline tumour site	Progression tumour site	Treatment during interim
MR1002	L trunk, posterior	L trunk, posterior	Vemurafenib
MR1004	L posterior ear	L supraclavicular fossa	Ipilimumab, Pazopanib
MR1006	R thigh	R thigh	Vemurafenib, Pan-RAF inhibitor (MLN2480)
MR1010	R trunk, anterior	L trunk, posterior	Vemurafenib
MR1012	L external iliac lymph node	L external iliac lymph node	Dabrafenib/trametinib
MR1014	R forearm	R cheek	Vemurafenib
MR1020	L axilla	L axilla	Vemurafenib
MR1022	R chest wall	R chest wall	Dabrafenib/trametinib
MR1032	L upper limb, anterior	L upper limb, anterior	Pan RAF Inhibitor (MLN2480)

Table 3.1 **Tumour sites and treatments for each patient**
For each of the 9 patients in this cohort, their baseline and progression tumour biopsy sites, and treatments received, are shown.

3.7.2 Sample processing

Blood samples were collected in S-Monovette 9mL EDTA tubes and inverted 8-10 times. For plasma collection, samples were centrifuged at 1600g for 10 minutes within an hour of the blood draw, and then an additional centrifugation of 20,000g for 10 minutes was carried out. Plasma samples were stored at -80°C.

3.7.3 Fresh frozen tumour, buffy coat and plasma DNA extraction

Up to 30mg of each fresh frozen tissue biopsy sample was combined with 600µL RLT buffer (Qiagen), then placed in a Precellys CD14 tube (Bertin Technologies) and homogenised at 6500 rpm for two bursts of 20 seconds separated by 5 seconds. DNA was extracted using the AllPrep extraction kit (Qiagen) as per the manufacturer's protocol. Genomic DNA was extracted from 10 mL whole blood using the Gentra Puregene Blood Kit (Qiagen) as per the manufacturer's protocol. Cell-free DNA was extracted from 2mL plasma using the QIAsymphony (Qiagen) according to the manufacturer's protocol.

Digital PCR was carried out using a Fluidigm Biomark^{TM} HD, run with 55 cycles of PCR using a hot-start polymerase. To quantify the cfDNA concentration of each sample, digital PCR was carried out using Taq-man probes for the housekeeping gene RPP30, and XenT, labelled with ROX and FAM, respectively.

3.7.4 Exome sequencing

Library preparation of 5µg genomic DNA, and sequencing and variant calling were performed as described by Varela et al. [274], using the Agilent SureSelectXT Human All Exon 50Mb bait set. Eight samples were multiplexed per pool and each pool loaded on to two lanes of an Illumina HiSeq 2000, giving an average 8GB of unique mapped reads per sample with an average of 80% of base pairs covered by >20 reads.

3.7.5 Amplicon sequencing

Tagged amplicon sequencing (TAm-Seq) was performed as previously described [29]. Panels were designed based on point mutations identified in either the baseline or progression tissue biopsies that had a depth of 10 reads or more. For each patient, a mean of 40 mutations (range 33-55) were selected for TAm-Seq. BRAF and *NRAS* hotspot mutations were included in each panel where they were detected in that patient, plus additional shared and private

mutations spanning a range of allele fractions. Samples were sequenced in duplicate to a mean depth of 32,000x on a HiSeq 2500. A depth-weighted average mutant allele fraction was calculated for each data point.

3.7.6 Shallow Whole Genome Sequencing

Libraries generated using the Rubicon ThruPLEX protocol were pooled at equimolar concentration then sequenced on a HiSeq 4000 to achieve a target coverage of 0.5-1x per library. Reads were de-multiplexed, mapped, and primer-dimers were removed *in silico*. Copy number was estimated using HMMcopy with the default settings [271], which first counts the number of reads in bins of equal genomic length, corrects based on local GC content, segments and classifies by copy number profile using a Hidden Markov Model.

3.7.7 Dilution series

To assess the limit of detection of TAm-Seq with a patient-specific approach, we generated a serial dilution of a sample with a high level of ctDNA, serially diluted in extracted cfDNA from a healthy individual (Seralabs). Extracted plasma cfDNA from patient MR1022 at their final time point was used for this dilution series. The concentration of each of the eluates was determined using digital PCR, then the samples were equalised in concentration using water. The patient sample was diluted 5-fold serially for a total of 6 dilutions (5x to 15,625x dilutions).

3.8 Bioinformatics and Statistical Methods

3.8.1 Mutation detection

To determine whether mutations are significantly above background noise, detection was performed per locus against a panel of 31 healthy control samples, as follows:

- At each interrogated locus, the background mutant allele fraction (AF) was determined as the depth-weighted mean AF in control samples.
- For each primer pair, the mean + 4 standard deviation threshold was calculated in a panel of control samples to give an expected false positive rate of 0.00006. Plasma samples were classified as positive if any primer pairs crossed this threshold.
- When these thresholds were applied to the same control samples, a false-positive rate of 0.0066 was observed.

3.8.2 Calculation of mean allele fraction

After loci were classified as detected/non-detected, non-detected loci were set as having an allele fraction of zero for the calculation of mean allele fraction, then a depth-weighted mean allele fraction was calculated across all patient-specific loci.

3.8.3 Receiver Operating Curve (ROC) analysis

easyROC [275] was used to fit ROC curves and determine the Area Under the Curve (AUC) and standard error of the AUC.

3.9 Supplementary Figures

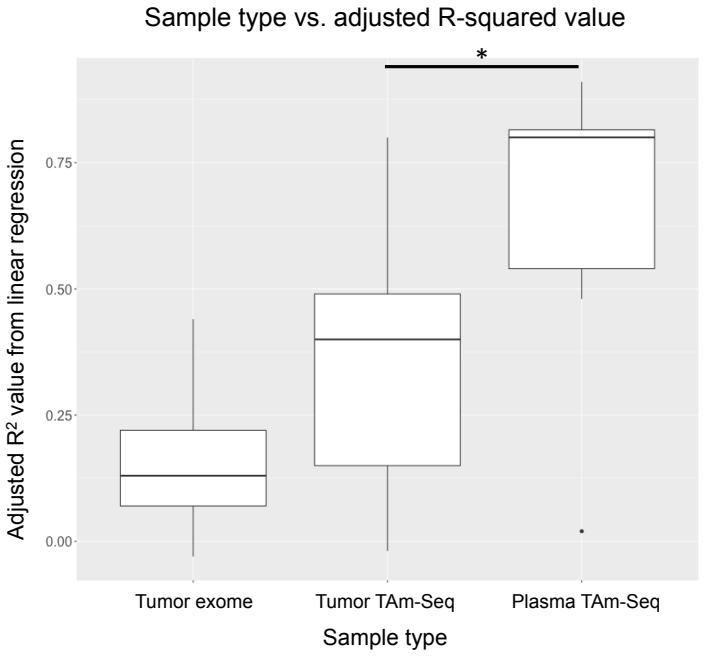


Fig. 3.9 **Box plots of R^2 values between two samples sequenced with either TAm-Seq or exome sequencing**
Box plots show the adjusted R^2 values based on the linear model fitted to the baseline vs. progression mutant allele fraction data from Figures S1-S9. Two patients were excluded from this analysis because their median plasma ctDNA mutant allele fraction was below 0.5% (using amplicon sequencing). Between baseline and progression time points, mutant allele fractions in plasma showed a significantly higher R^2 value than tumour sequencing ($P<0.05$; one-sided Wilcoxon rank sum test).

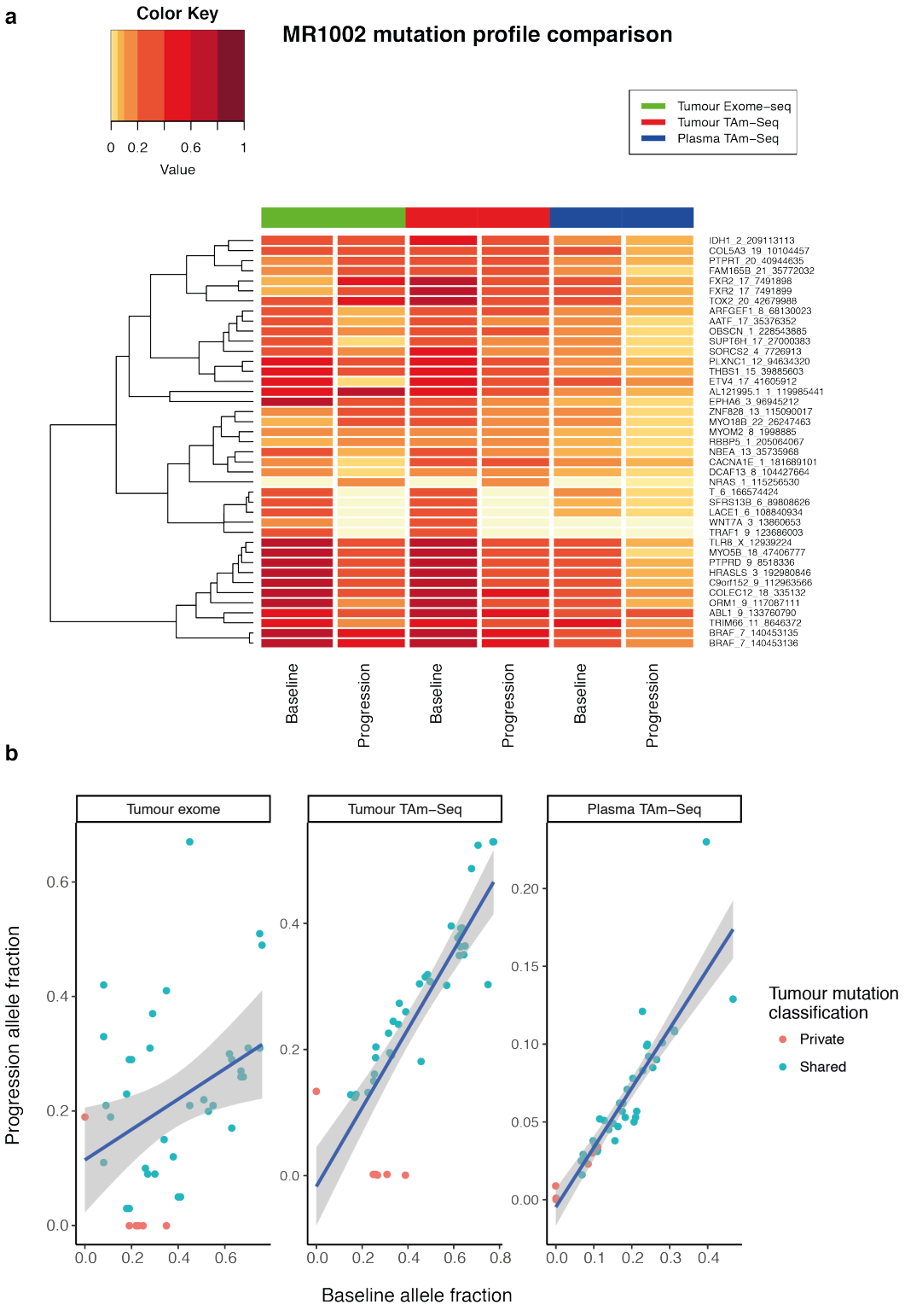


Fig. 3.10 **Tumour and plasma mutation profiles - MR1002**
Panel A shows a heat map of mutant allele fractions for each patient-specific locus at matched time points. Panel B compares mutant allele fractions between time points for each method. Panel C shows a matrix of pairwise comparisons of mutant allele fraction. In panels B and C, tumour mutations were classified as shared if they were detected in both tumour samples (blue), or private if detected in only one (red), based on amplicon sequencing.

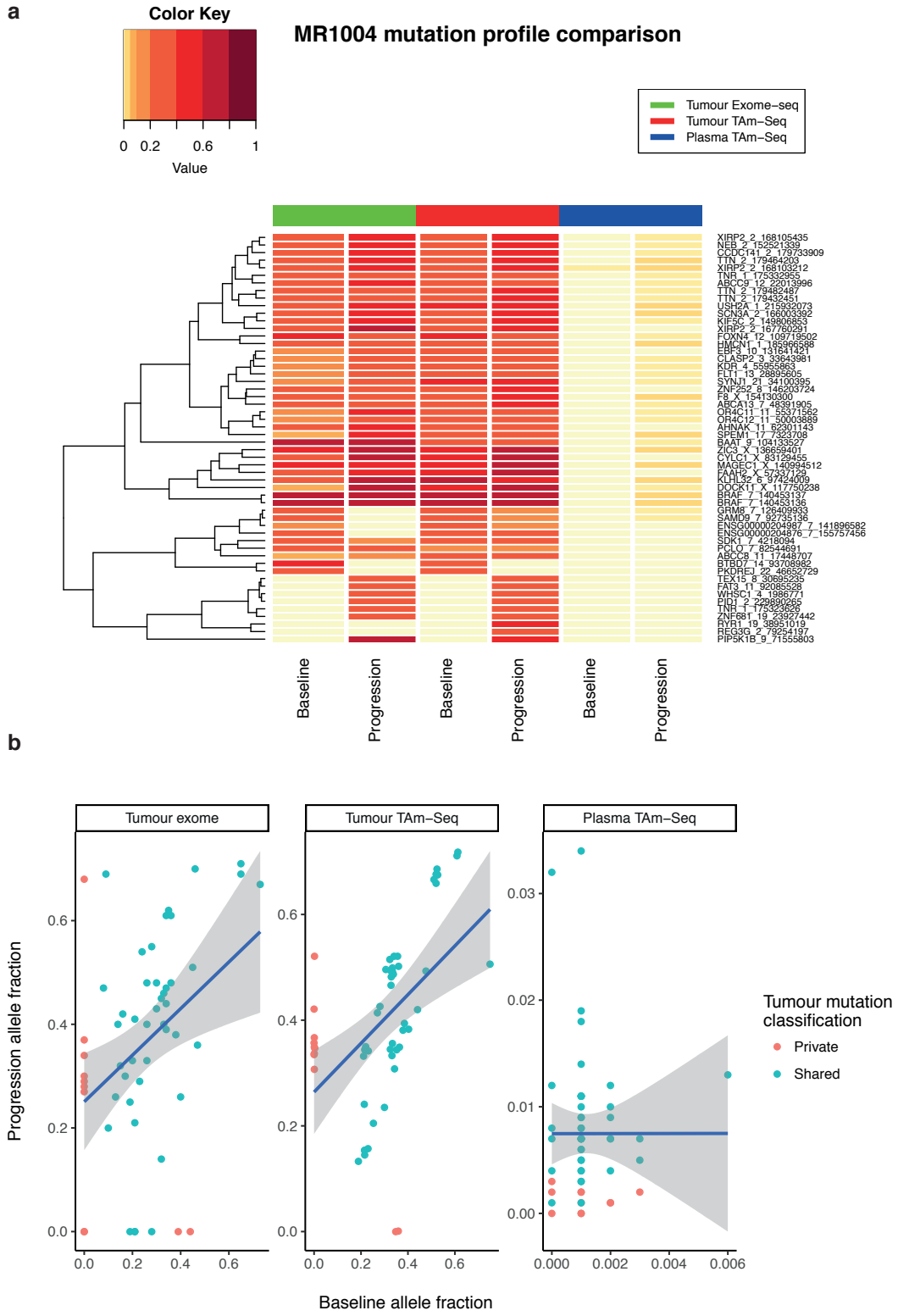


Fig. 3.11 **Tumour and plasma mutation profiles - MR1004**
Panel A shows a heat map of mutant allele fractions for each patient-specific locus at matched time points. Panel B compares mutant allele fractions between time points for each method. Panel C shows a matrix of pairwise comparisons of mutant allele fraction. In panels B and C, tumour mutations were classified as shared if they were detected in both tumour samples (blue), or private if detected in only one (red), based on amplicon sequencing.

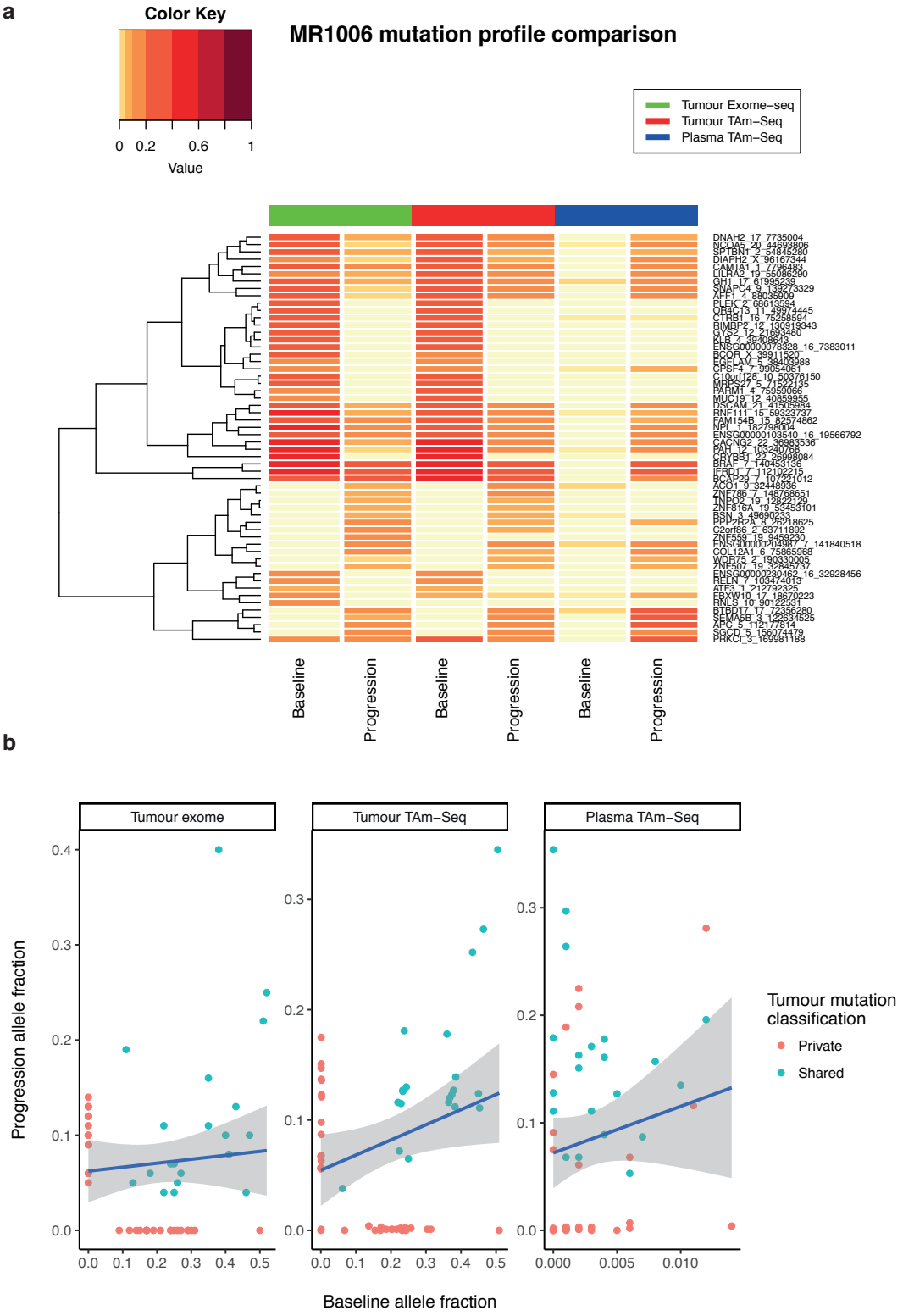


Fig. 3.12 **Tumour and plasma mutation profiles - MR1006**
Panel A shows a heat map of mutant allele fractions for each patient-specific locus at matched time points. Panel B compares mutant allele fractions between time points for each method. Panel C shows a matrix of pairwise comparisons of mutant allele fraction. In panels B and C, tumour mutations were classified as shared if they were detected in both tumour samples (blue), or private if detected in only one (red), based on amplicon sequencing.

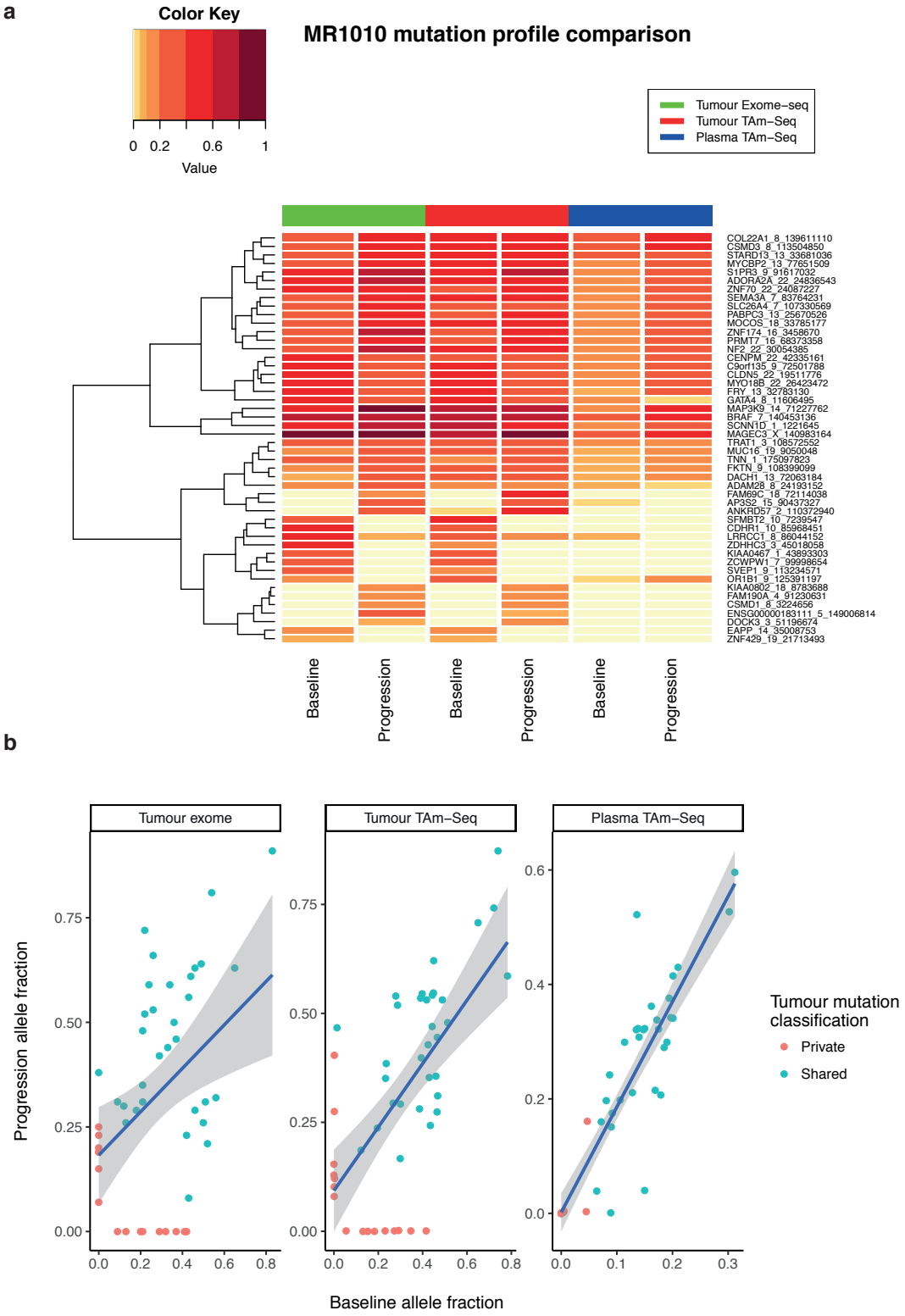


Fig. 3.13 **Tumour and plasma mutation profiles - MR1010**
Panel A shows a heat map of mutant allele fractions for each patient-specific locus at matched time points. Panel B compares mutant allele fractions between time points for each method. Panel C shows a matrix of pairwise comparisons of mutant allele fraction. In panels B and C, tumour mutations were classified as shared if they were detected in both tumour samples (blue), or private if detected in only one (red), based on amplicon sequencing.

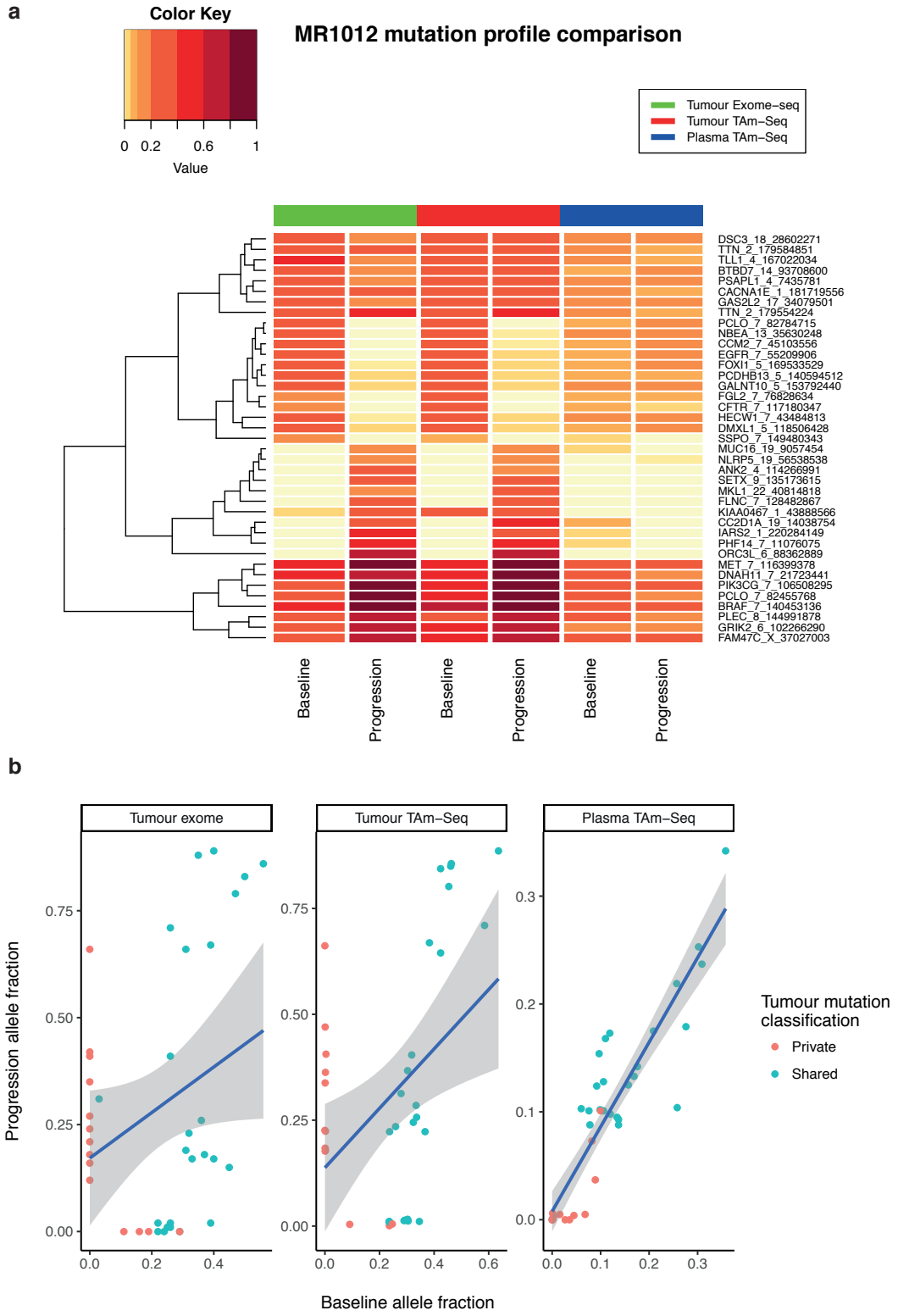


Fig. 3.14 **Tumour and plasma mutation profiles - MR1012**
Panel A shows a heat map of mutant allele fractions for each patient-specific locus at matched time points. Panel B compares mutant allele fractions between time points for each method. Panel C shows a matrix of pairwise comparisons of mutant allele fraction. In panels B and C, tumour mutations were classified as shared if they were detected in both tumour samples (blue), or private if detected in only one (red), based on amplicon sequencing.

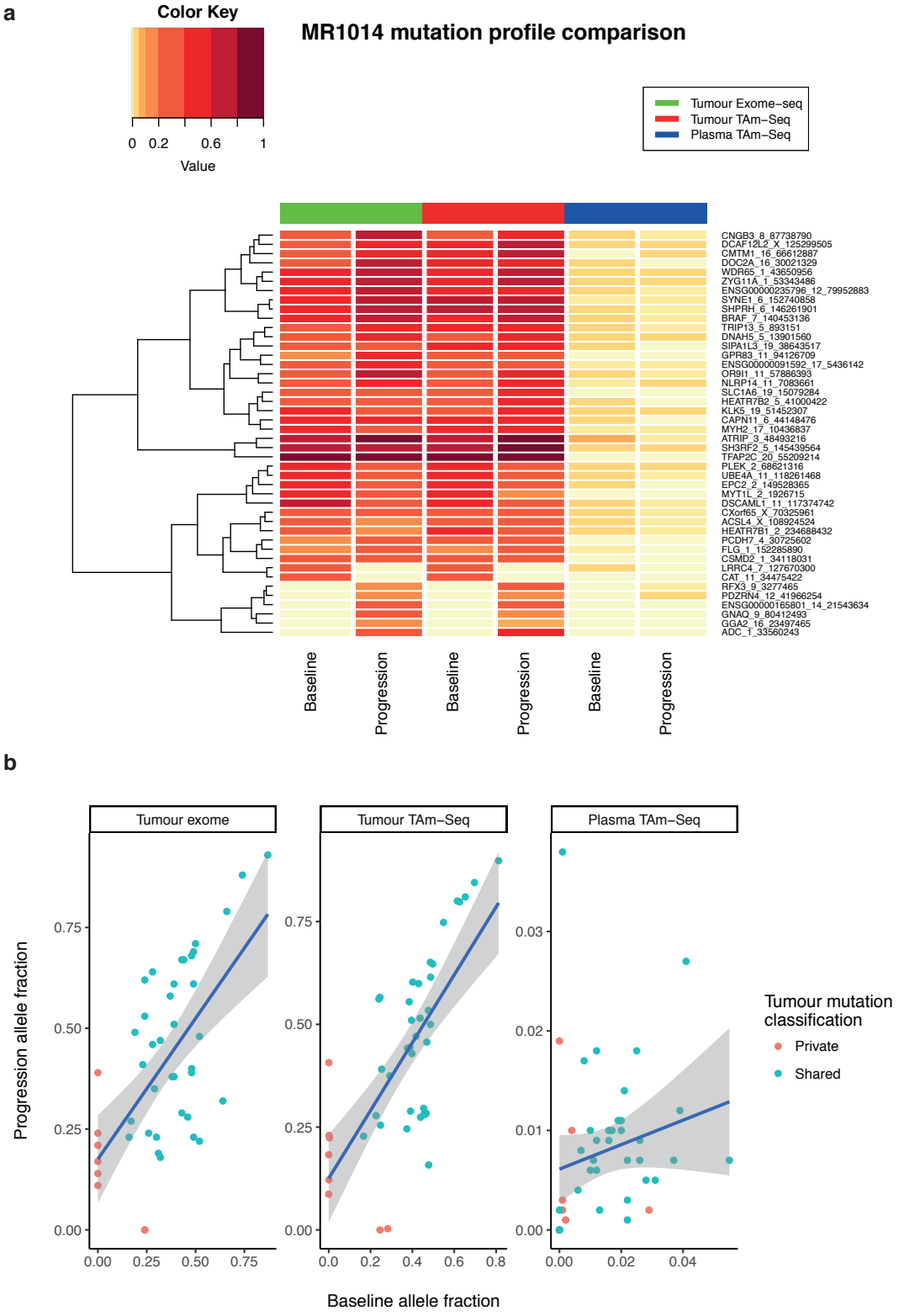


Fig. 3.15 **Tumour and plasma mutation profiles - MR1014**
Panel A shows a heat map of mutant allele fractions for each patient-specific locus at matched time points. Panel B compares mutant allele fractions between time points for each method. Panel C shows a matrix of pairwise comparisons of mutant allele fraction. In panels B and C, tumour mutations were classified as shared if they were detected in both tumour samples (blue), or private if detected in only one (red), based on amplicon sequencing.

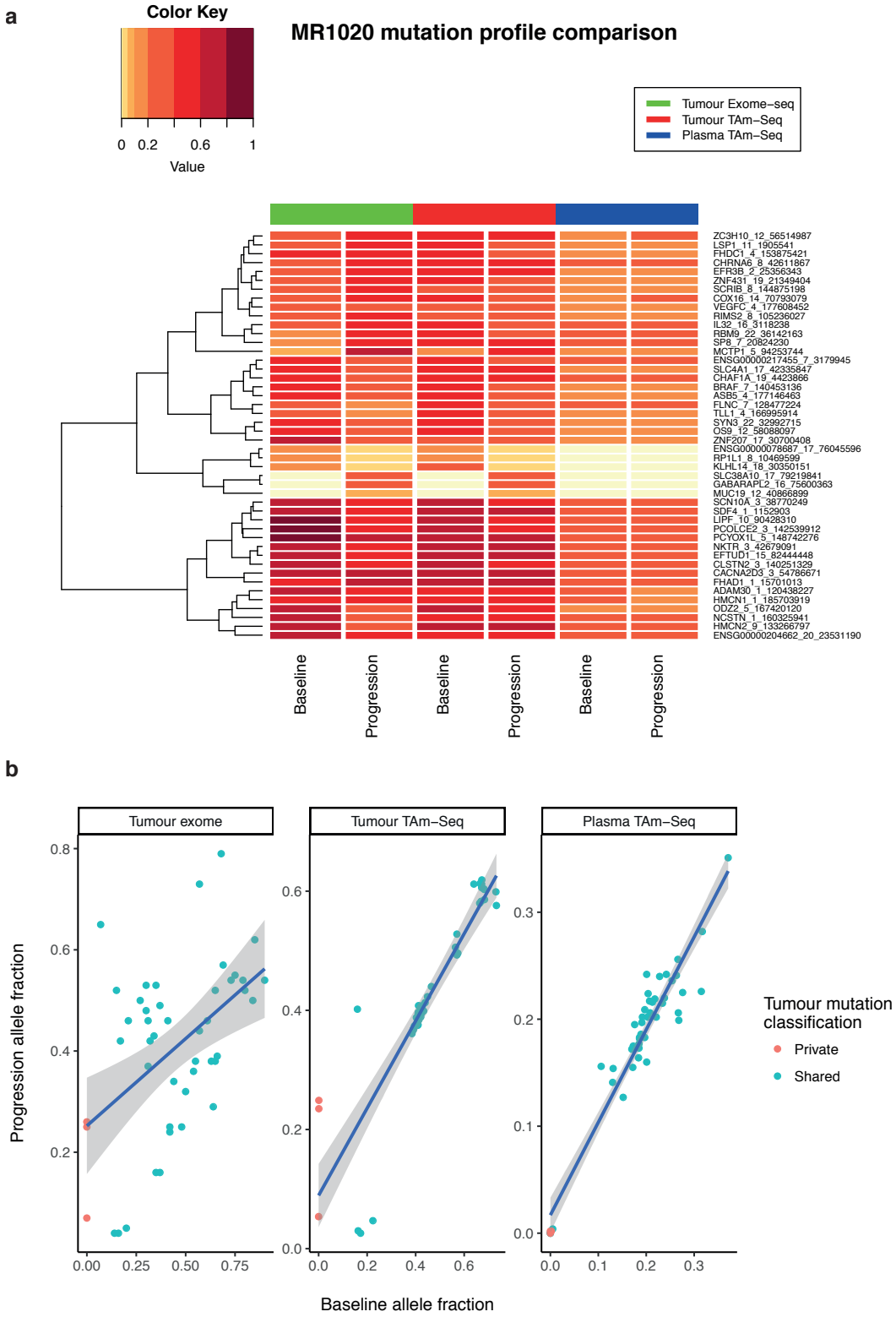


Fig. 3.16 **Tumour and plasma mutation profiles - MR1020**
Panel A shows a heat map of mutant allele fractions for each patient-specific locus at matched time points. Panel B compares mutant allele fractions between time points for each method. Panel C shows a matrix of pairwise comparisons of mutant allele fraction. In panels B and C, tumour mutations were classified as shared if they were detected in both tumour samples (blue), or private if detected in only one (red), based on amplicon sequencing.

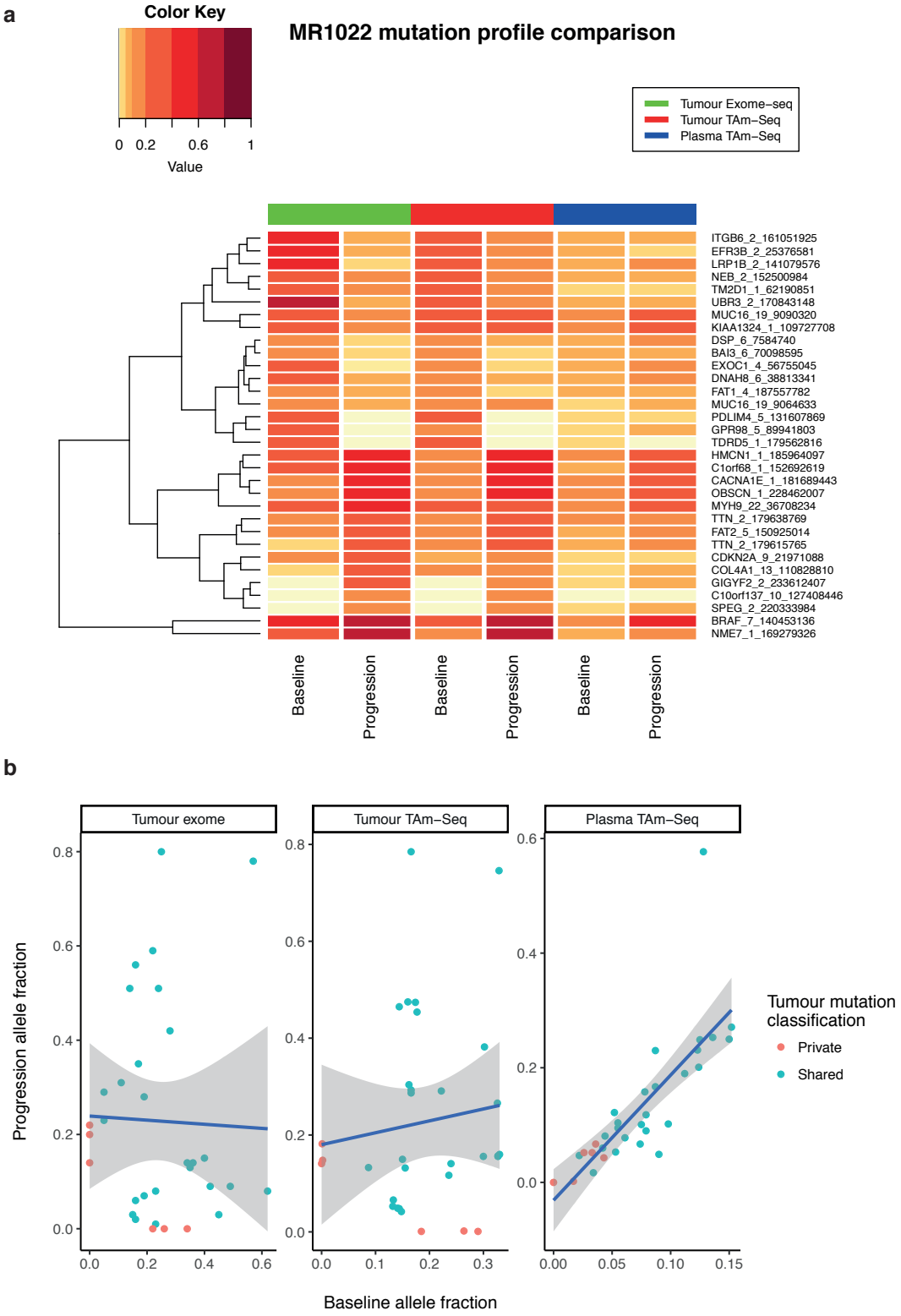


Fig. 3.17 **Tumour and plasma mutation profiles - MR1022**
Panel A shows a heat map of mutant allele fractions for each patient-specific locus at matched time points. Panel B compares mutant allele fractions between time points for each method. Panel C shows a matrix of pairwise comparisons of mutant allele fraction. In panels B and C, tumour mutations were classified as shared if they were detected in both tumour samples (blue), or private if detected in only one (red), based on amplicon sequencing.

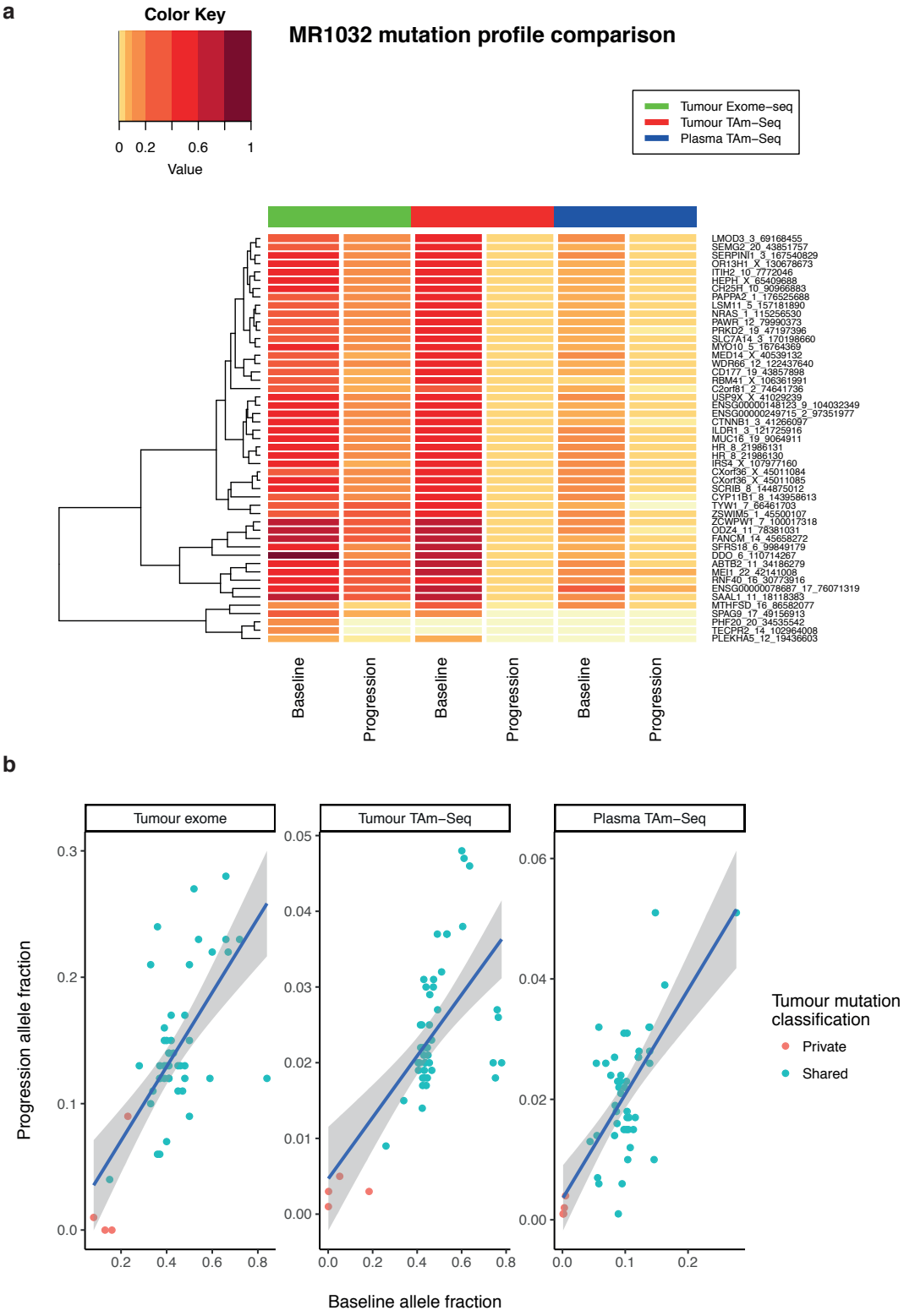


Fig. 3.18 **Tumour and plasma mutation profiles - MR1032**
Panel A shows a heat map of mutant allele fractions for each patient-specific locus at matched time points. Panel B compares mutant allele fractions between time points for each method. Panel C shows a matrix of pairwise comparisons of mutant allele fraction. In panels B and C, tumour mutations were classified as shared if they were detected in both tumour samples (blue), or private if detected in only one (red), based on amplicon sequencing.

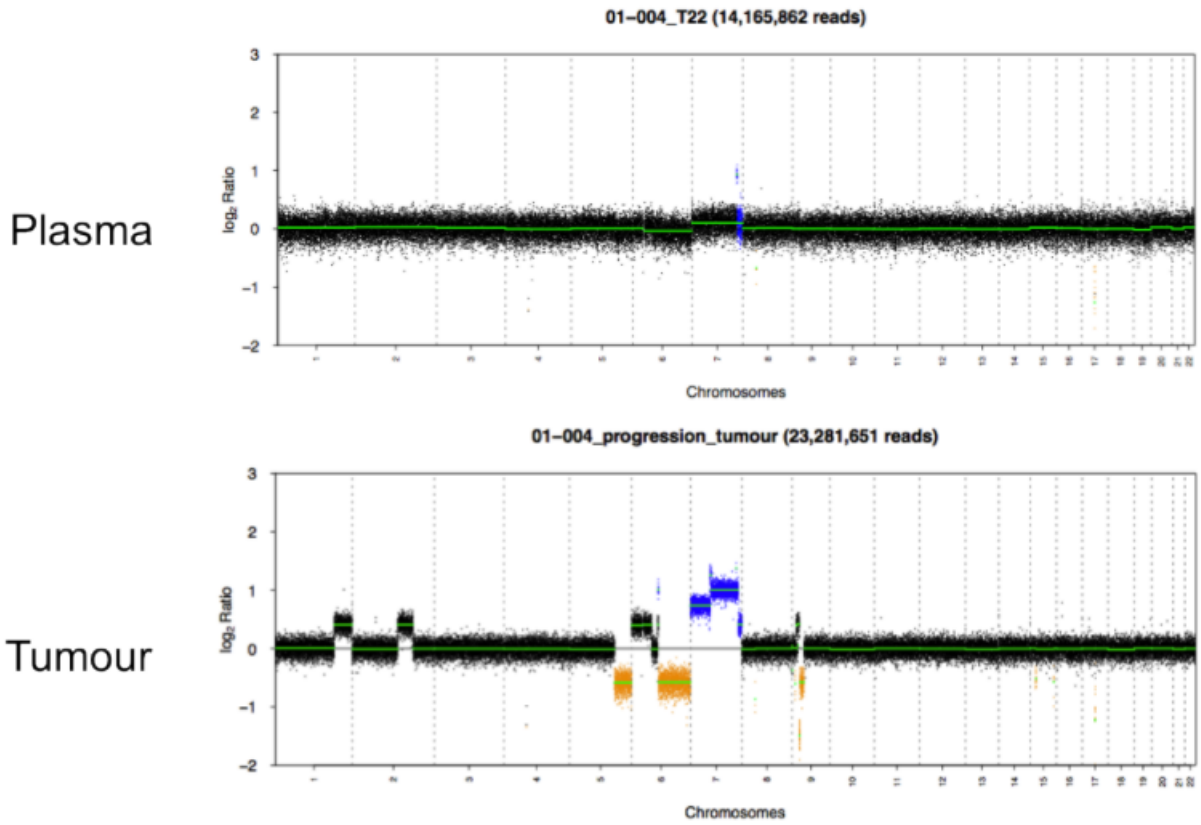


Fig. 3.19 **BRAF amplification in tumour and plasma - patient MR1004**
Shallow WGS was used to assess copy number profile for patient MR1004 in both tumour and plasma samples. Log_2 ratios are shown for each bin. Bins were segmented and called as described in Method 3.7.6. Blue indicates a called amplified segment, and yellow indicates a deleted segment.

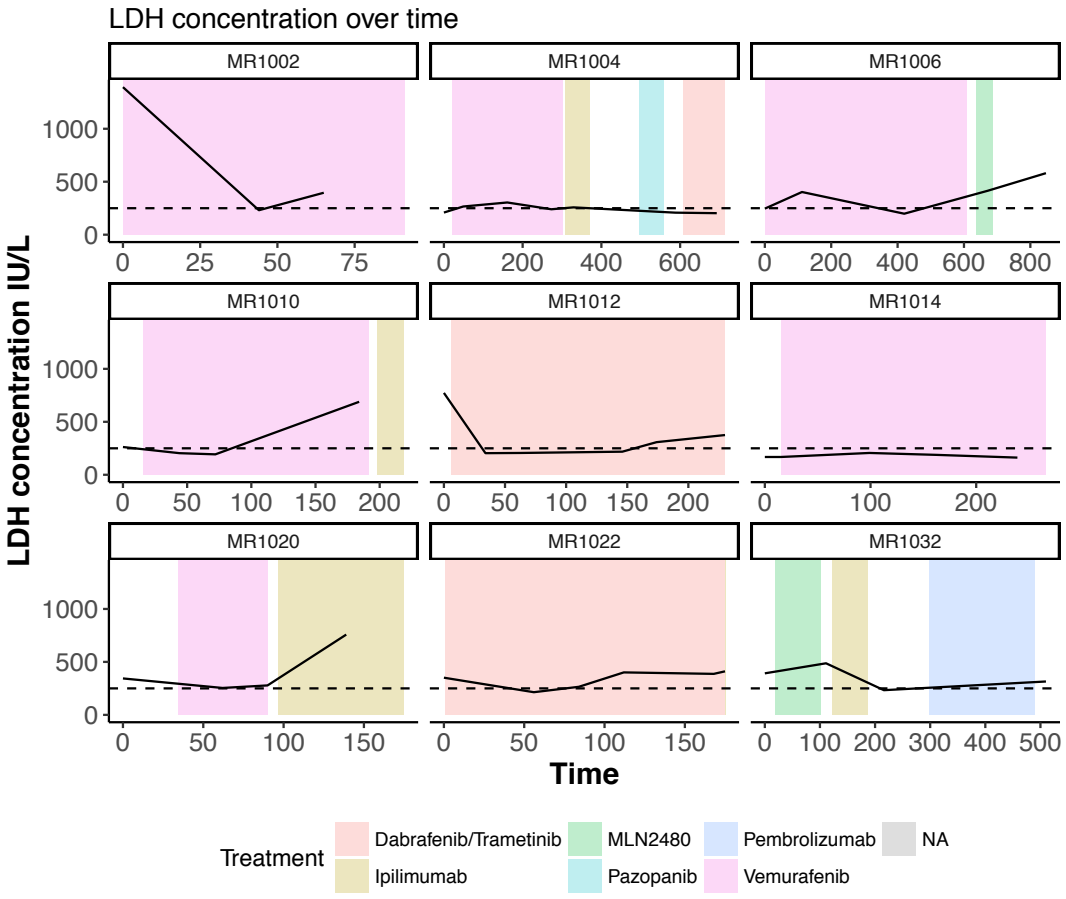


Fig. 3.20 **Serum lactate dehydrogenase concentration over time**
Serum LDH concentrations (IU/L) are plotted over time, and treatments are indicated by shaded boxes. The upper limit of normal (250 IU/L) is shown with a horizontal line.

Chapter 4

Monitoring ctDNA to parts per million by integration of variant reads

4.1 Attribution

I developed the INVAR pipeline with Katrin Heider; all decisions on Figures included in the manuscript were agreed by both Katrin and I following discussion, then approved by Nitzan. This chapter is adapted from a manuscript we resubmitted in July 2018 (full citation in Section 6.2).

I should emphasise that the INVAR project certainly was a two-person job (at least), as it required us to carry out wet lab, dry lab, pipeline development tasks, and paper writing - often in parallel. Our joint first co-authorship on the INVAR manuscript reflects this, and I trust that the university will allow Katrin to use her fair share of our manuscript in her thesis. I will detail our specific intellectual and experimental contributions below.

First, to summarise our contributions to this pipeline, the number of commits per person to the Bitbucket repository are shown in Table 4.1. Once the data were processed with this pipeline, Katrin and I proceeded to explore the data. The commits for this exploratory work, data visualisation and summarisation for the paper are shown in Table 4.2.

Jonathan Wan	901
Katrin Heider	329
Eyal Fisher	32
James Morris	7

Table 4.1 **Bitbucket commits for the INVAR pipeline**
Table of commits to the INVAR Bitbucket repository. Katrin Heider is another PhD student in the lab (2015-19), working on custom capture sequencing for NSCLC. Eyal Fisher is a mathematics MPhil student who carried out a 3-month rotation in our lab (during Spring 2018), who we consulted with to develop methods that we implemented in the pipeline. James Morris is a bioinformatician in the lab who mentored us from a bioinformatics standpoint throughout, and helped us troubleshoot errors in the pipeline.

Jonathan Wan	316
Katrin Heider	152

Table 4.2 **Bitbucket commits for data exploration**
Commits to the repository where we plotted all the figures for the INVAR paper. We jointly ran all samples through the pipeline, and following data being output, we visualised the data in this repository.

4.1.1 Methods line-by-line contribution

- **Methods 4.6.1, 4.6.2 – Patient cohort**. Sample and clinical information collection and plasma sample processing were performed by the respective study co-ordination teams.

- **Methods 4.6.3 - 4.6.5 – Tumour and plasma DNA extraction, quantification and library preparation**. I performed these wet lab steps for the MELR (stage IV melanoma) and AVAST-M (stage II-III resected melanoma) tumour and plasma samples, and Katrin did the same for the LUCID (stage I-IIIA NSCLC pre-surgery) samples in parallel. These steps were performed with the aim of generated individualised capture sequencing data for each of our cohorts, suggested by Nitzan.

- **Method 4.6.6 – Custom hybrid-capture panel design and sequencing of plasma**. Katrin and I independently generated custom hybrid-capture baitsets for each of our cohorts, again suggested by Nitzan.

- **Methods 4.6.7, 4.6.8 – Exome and WGS of plasma with INVAR**. After we demonstrated that the INVAR method could be applied to custom capture sequencing data, Katrin and I agreed that it in theory could be run on any deep sequencing data with patient-specific mutations. Thus, I explored its use on exome and WGS data from MELR plasma samples using plasma WES/WGS data I had previously generated. To confirm the sensitivity of this approach, Katrin did the wet lab work to generate additional plasma exome sequencing libraries.

- **Method 4.6.9 – Experimental spike-in dilution series**. I generated a spike-in dilution series to test the sensitivity of using thousands of mutations as opposed to 50 mutations which I had previously tested in the individualised TAm-Seq work. I applied the sequencing panel designed in Method 4.6.6 to these dilution series.

- **Method 4.6.10 – Imaging**. Andrew Gill and Ferdia Gallagher analysed volumetric CT imaging at time points matching with plasma sample data.

- **Method 4.7.1 – Generation of patient-specific mutation lists**. Tumour mutation calling was performed using our in house pipeline, developed by our Bioinformatics core. I ran the pipeline with some slight modifications for MELR and AVAST-M. Katrin ran the pipeline on LUCID samples. We agreed on the filters jointly through discussion and exploration of the data.

- **Method 4.7.2 – Data processing and error-suppression**. I modified our sequencing data pipeline to take ThruPLEX Tag-Seq data, which has specific requirements for pre-processing prior to read collapsing. After comparing error rates and proportions of data retained, we opted for a family size of 2 (Fig. 4.28), similar to Newman et al. [111]. Katrin and I jointly ran samples through the pipeline, and Katrin wrote scripts to run multiple cohorts of samples at once.

- **Method 4.7.2 – Concatenation of data across time points**. Nitzan suggested the possibility of concatenating sequencing data to leverage signal across multiple time points, which I implemented with the code shown.

- **Method 4.7.3 – Data filtering**. I suggested and implemented these filters. In addition, instead of assessing background error rates by mutation class, I increased the resolution into error rates by annotating each mutation with its trinucleotide context. Katrin and I jointly explored this data on our respective cohorts and jointly reached the conclusion that error rates should be evaluated by trinucleotide context. We changed the threshold for identifying a noisy locus from $\geq$3 control samples to requiring signal in at least 10% of control samples.

- **Method 4.7.4 – Feature annotation**. Trinucleotide context-based error rates were favoured over mutation class error rate, as I showed, and Katrin confirmed, variability in error rate between different trinucleotide contexts within the same class. (Fig. 4.6). For annotation with fragment size, Dineika Chandrananda had previously written a script to obtain the fragment length of reads from a BAM file, which I adapted with assistance from James Morris and incorporated in the pipeline.

- **Method 4.7.5 – Patient-specific outlier suppression**. I conceived and developed this approach to identify ctDNA signal at a locus that is not consistent with all other patient-specific loci. Katrin confirmed the utility of the approach on her cohort.

- **Method 4.7.6 – Statistical model for detection, estimation of read length distribution and calculation of IMAF**. All mathematical equations and their definitions were written by Eyal Fisher (Mathematics MPhil rotation student) based on our discussions of the problem and my previous approaches (*data not shown*). I wrote the explanatory text around the equations. We worked together to integrate his code into the pipeline, and I explored the resulting data. Katrin and I performed troubleshooting of the code.

1771 – **Tumour allele fraction weighting**. Katrin and I independently demonstrated
1772 the potential utility of tumour allele fraction weighting from exploratory data
1773 analysis in each of our cohorts. These findings were consistent with previous
1774 studies on tumour heterogeneity [124]. We agreed that we should incorporate
1775 this into the method, and so I outlined the problem to Eyal. Thus, Eyal wrote
1776 an equation to incorporate tumour allele fraction (AF_i) when calculating INVAR
1777 likelihood ratios.

1778 – **Fragment size weighting**. The Mouliere et al. [276] paper demonstrated the
1779 potential sensitivity benefit of leveraging fragment sizes of ctDNA. However,
1780 size-selection can cause drop out of rare mutant alleles when ctDNA levels are
1781 very low, as is the case here. Therefore, Katrin and I discussed and agreed that
1782 a size-weighting-based approach would be appropriate to boost signal where
1783 possible, but retain all signal. We proposed this extension of the inclusion of size
1784 to Eyal. I updated the pipeline's code so that the format of the data would be
1785 compatible with his size-weighting script.

1786 – **GLRT**. Eyal selected an appropriate test to generate an output value based on
1787 exploratory analyses on our samples and on in silico data.

1788 • **Method 4.7.7 - Estimation of the read length distribution**. In order to perform
1789 weighting by size, a reference size distribution is needed. Eyal highlighted the need
1790 for this, and the need to smooth the distribution. I generated the data in an appropriate
1791 format, and tested the effect of different levels of smoothing (Fig. 4.35).

1792 • **Method 4.7.8 – Calculation of IMAF**. All equations and definitions in this section
1793 were written by Eyal. I wrote the explanatory text following discussion with Eyal.

1794 • **Method 4.7.9 – Detection based on classification of INVAR scores**. I wrote the
1795 code in the pipeline to apply a ROC-based cut-off for INVAR score.

- **Method 4.7.10 – Calculation of informative reads**. The concept of 'informative reads' as a quantifier of sensitivity based on the total number of patient-specific reads on target was suggested by Nitzan.

- **Method 4.7.11 – Survival analysis**. Kaplan-Meier analysis was performed by our trial statistician (Andrea Marshall). We were blinded to patient outcomes. I provided Andrea with ctDNA values and detection status for all samples in order for her to do this.

- **Method 4.7.12 – Estimated detection rates with fewer IR**. I outlined the problem to Eyal and he kindly provided us with an equation to calculate the 95% confidence interval of a sample's allele fraction given that zero mutant reads were detected in n IR.

4.2 Summary

Circulating tumour DNA (ctDNA) is increasingly used to monitor tumour responses [59, 151]. In patients with low disease burden, ctDNA detection rates are low due to the presence of few or no copies of any individual mutation in each sample [129, 117]. Sensitivity may be increased by collecting larger plasma volumes, but this is not feasible in practice. Here we demonstrate that sensitivity can be greatly enhanced by analysing a large number of mutations via sequencing. Although cancers have thousands of mutations in their genome [234], previously published work measured only individual or up to 32 tumour-specific mutations in plasma [29, 59, 106, 117]. We sequenced in plasma 10^2-10^4 mutated loci per patient, using custom capture panels, whole exome or shallow (low-coverage) whole genome sequencing (WES/WGS). We developed an analytical method for INtegration of VAriant Reads (INVAR) that aggregates reads carrying tumour mutations, and uses a statistical model to assign confidence to error-suppressed reads based on mutation context, fragment length and tumour representation. We applied this algorithm to 143 plasma samples from 47 patients with stage II-IV melanoma, 19 patients with stage I-IIIA non-small cell lung cancer (NSCLC), and 26 healthy individuals. ctDNA was detected to 1 mutant molecule per million, and tumour volumes of ~1cm^3. INVAR detected ctDNA in 66% of patients with stage I-II NSCLC. In patients with stage II-III melanoma who relapsed after resection, ctDNA was detected within 6 months of surgery and prior to relapse in 50% of cases. INVAR can enhance detection of ctDNA in samples with limited input or sequencing coverage. Using WGS (0.6x) and WES data, ctDNA was detected to concentrations of 10^{-4} by integrating data from thousands of mutant loci. As tumour sequencing becomes more widespread, this method has potential to enhance detection of minimal residual disease, and to enable routine cancer monitoring using low-depth sequencing and from low-volume blood samples, which might be self-collected.

4.3 Introduction

In patients with early-stage cancer [129], or patients at all stages undergoing treatment [59], ctDNA concentrations can be low and may often result in false-negatives due to sampling error, even for an assay with perfect analytical performance [59] (Fig.4.1). Sequencing errors can further limit detection. To improve sensitivity, studies have analysed larger volumes of plasma from multiple blood tubes [170, 185], and/or used sequencing panels covering 18-507 genes (2 kb – 2 Mb of the genome) [117, 170, 203, 277, 278]. Analysis in plasma of up to 32 patient-specific mutations (identified beforehand via tissue analysis) achieved detection to levels of 1 mutant molecule per 25,000 copies in a patient with NSCLC [106]. ctDNA was detected in <50% for patients with stage I NSCLC [117, 203] and in only 13% of lung adenocarcinoma cases [117]. In early-stage patients who underwent surgery and later relapsed, ctDNA was detected after surgery in approximately 50% of breast or colorectal cancer patients [185, 186] and 15% of melanoma patients [266].

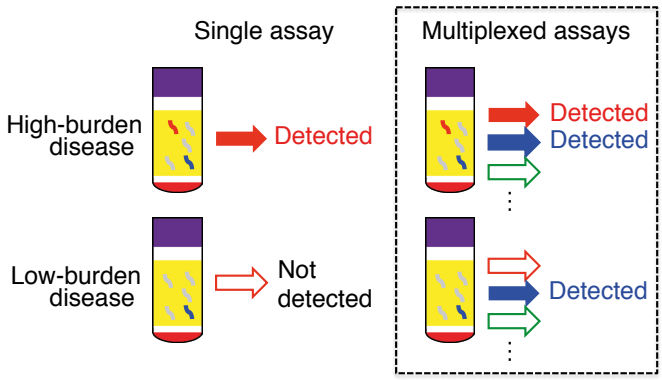


Fig. 4.1 **Targeting multiple mutations overcomes sampling error**
Even with perfect analytical performance, a single-locus assay can fail to detect low ctDNA levels due to random sampling of DNA fragments. This can be overcome by using multiplexed assays on the same sample.

To obtain information from a large number of mutations per patient, we sequenced plasma DNA using custom capture panels, WES or WGS. In analysing sequencing

data, ctDNA detection algorithms have previously relied on identification of individual mutations [117, 170, 203], which uses limited information inefficiently: any signal that does not pass a mutation calling threshold is discarded and lost. Studies have highlighted the potential advantages of aggregating signal across multiple loci to detect DNA from transplanted organs [7] or diluted tumour DNA [106], but this has not been applied to monitor cancer in plasma. To efficiently use sequencing information from plasma we developed INVAR. INVAR uses prior information from tumour sequencing to guide analysis and aggregate signal across 10^2-10^4 loci which are known to be mutated in the patient's cancer. It considers biological and technical features of ctDNA sequencing including trinucleotide error rate, ctDNA fragment length patterns and the allele fraction of each mutation in the patient's tumour (Fig. 4.2).

To identify patient-specific mutations, we performed tumour sequencing of 45 patients with stage II-IV melanoma and 19 patients with stage I-IIIA NSCLC. After mutation calling (Method 4.7.1), we generated patient-specific mutation lists (Supplementary Figs. 4.26, 4.27 and Table 4.3), consisting of a median of 625 mutations per patient with melanoma (IQR 411-1076) and 388 per patient with stage I-IIIA NSCLC (IQR 230–600). These lists were used as input for INVAR, together with plasma DNA sequencing data generated using custom capture sequencing panels (2,301x mean depth), WES (290x) and shallow WGS (sWGS, 0.6x, Fig. 4.3, Table 4.4).

The detection of ctDNA is limited by the amount of DNA analysed, which we evaluate as the number of haploid genomes analysed (hGA), equivalent to the average deduplicated sequencing coverage. In methods such as sWGS, DNA analysed is often <1 hGA (less than 1x coverage), and although this is often generated from nanogram amounts of DNA, it could in principle be generated from picograms of DNA. The number of Informative Reads (IR) covering tumour-mutated loci are equal (on average) to the number of hGA multiplied by the number of mutations analysed (Fig. 4.4, defined

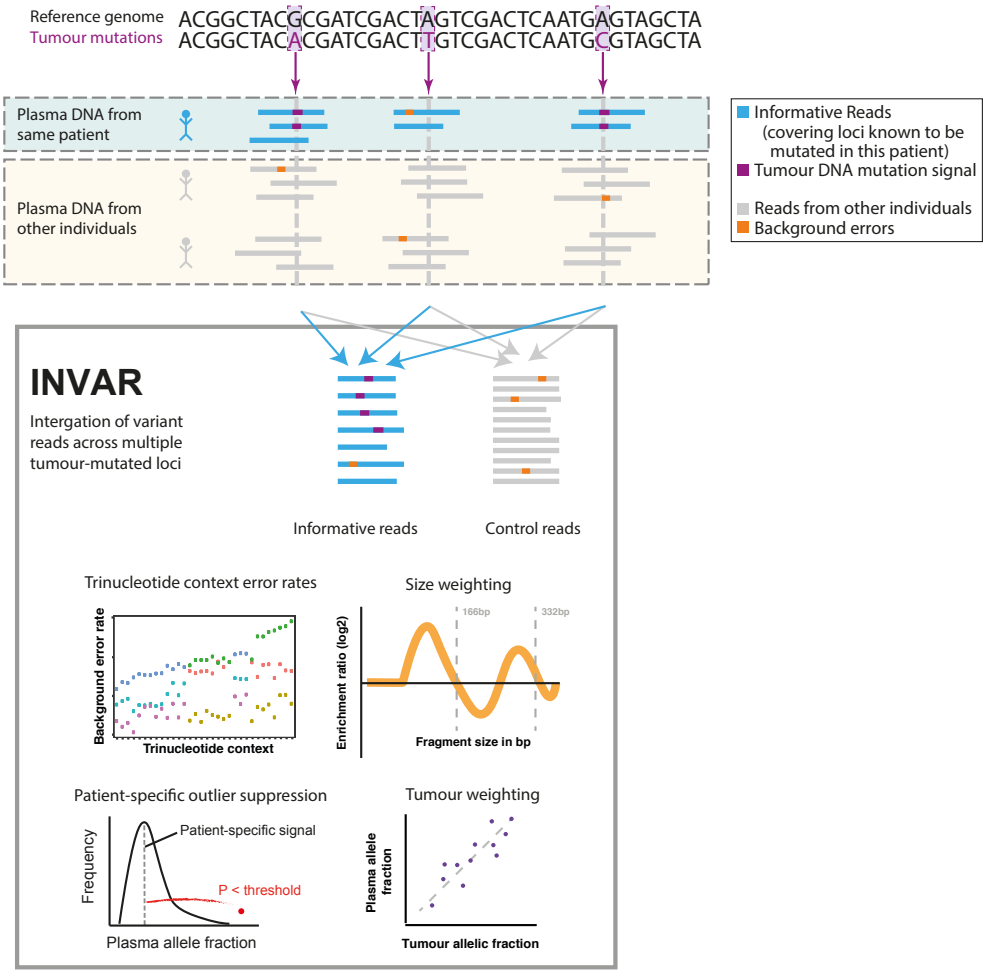


Fig. 4.2 **INtegration of VAriant Reads concept**
To overcome sampling error, signal was aggregated across hundreds to thousands of mutations. Here we classify samples (rather than mutations) as being significantly positive for ctDNA, or non-detected. Reads from a patient's sample that overlap loci in the patient-specific mutation list are indicated as 'patient-specific', whereas reads overlapping the same loci in other patients are indicated as 'non-patient-specific'. INVAR also incorporates additional sequencing information on fragment length and tumour allelic fraction, to enhance detection.

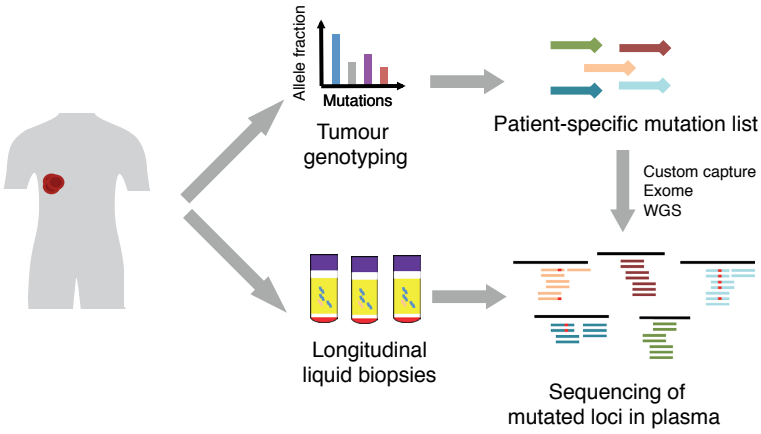


Fig. 4.3 **Experimental outline of study**
Tumour sequencing was performed to identify patient-specific mutations, which are targeted for high-depth sequencing data in matched plasma. In this study, custom capture sequencing, exome and whole-genome sequencing were used on plasma samples.

in Method 4.7.10). For example, 10^5 IR may be generated from 10,000 hGA and 10 mutated loci (deep sequencing of a narrow panel) or from 10 hGA across 10,000 loci (wide coverage of limited input). In a sample with ctDNA mutant allele fraction of 10^{-5}, observing a single mutant read across 10^5 IR would have a probability of 0.26 due to binomial sampling statistics, but this increases to 0.96 with 5 x 10^5 IR, highlighting the relationship between molecules sampled and the maximum sensitivity achievable.

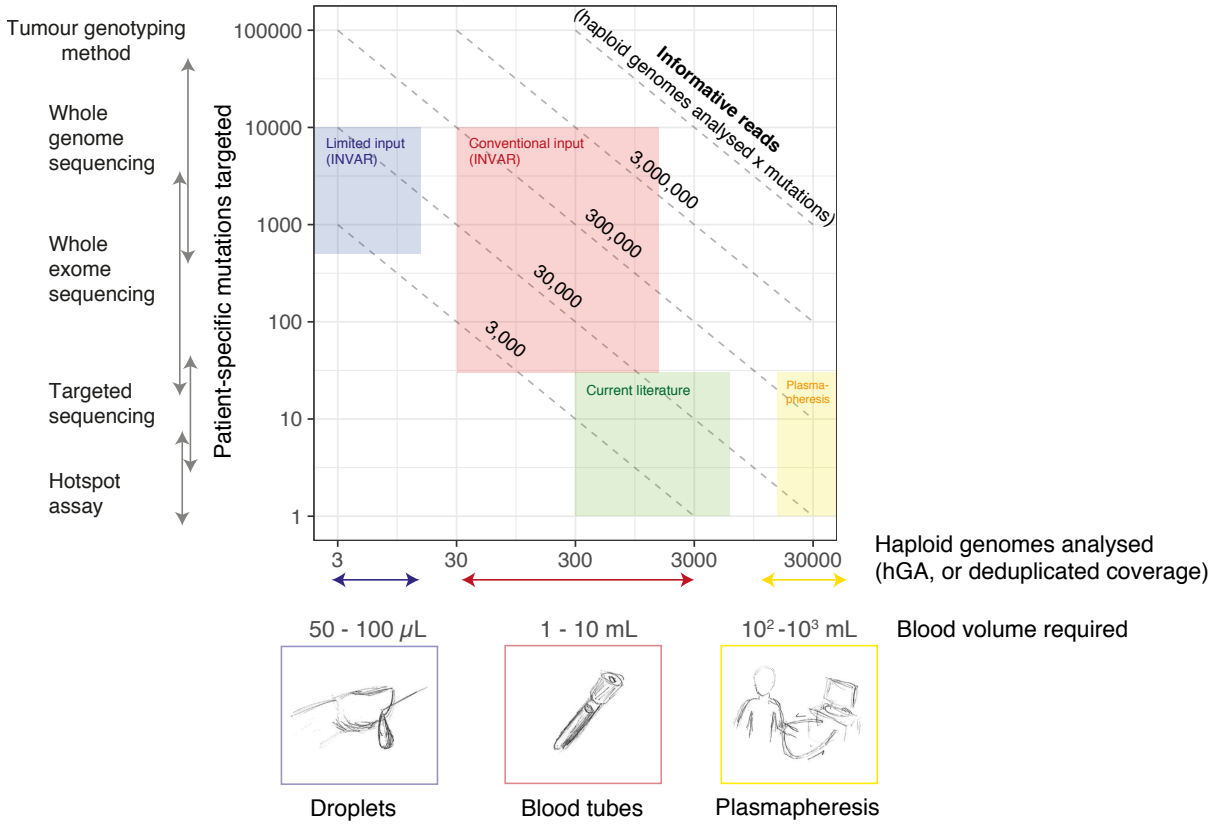


Fig. 4.4 **Working points for ctDNA analysis**
Illustration of the range of possible working points for ctDNA analysis using INVAR. Current methods often focus on analysis of ~10ng of DNA (300-3,000 haploid copies of the genome) across 1 to 30-50 mutations per patient. This results in ~10,000 informative reads (IR, equivalent to hGA x targeted loci), leading to frequently encountered detection limits of 0.01%-0.1% [203, 117]. Increasing the number of hGA and the number of mutations targeted result in a larger number of IR, and therefore, greater sensitivity.
We utilised integration of variant reads (INVAR) to explore the possible working points of sensitivity that can be achieved using 10^3-10^4 mutations. We applied this method to samples with both conventional levels of input mass (obtainable from millilitres of blood, red box) and from limited input (~10 hGA, blue shaded box). Theoretically, high sensitivity can be achieved through sampling larger volumes of blood, but often this may be unfeasible in cancer patients. The shaded boxes show different possible working points for ctDNA analysis. LOD, limit of detection; IR, informative reads; ng, nanogram; mL, millilitre.

4.4 Results

4.4.1 Development of the INVAR algorithm

Error-suppression

Increasing the number of informative reads addresses sampling error. To reduce the likelihood of false positive detection, the background error must be below the reciprocal of IR. As part of the INVAR workflow (Fig. 4.2), we reduced background error rates through the following: (i) by read-collapsing based on endogenous or exogenous unique molecular identifiers (UMIs) [112], (ii) excluding signal that was not supported by both forward and reverse reads, (iii) using a bespoke error model that assesses error rates for different mutation contexts from data outside patient-specific mutation lists, and (iv) excluding outlying signal from individual loci that were not consistent with the distribution of signals from other loci in that sample (Supplementary Fig. 4.32, Method 4.7.5). Together, these methods resulted in a combined reduction of background error by 131-fold on average across trinucleotide contexts (Figs. 4.5, 4.6).

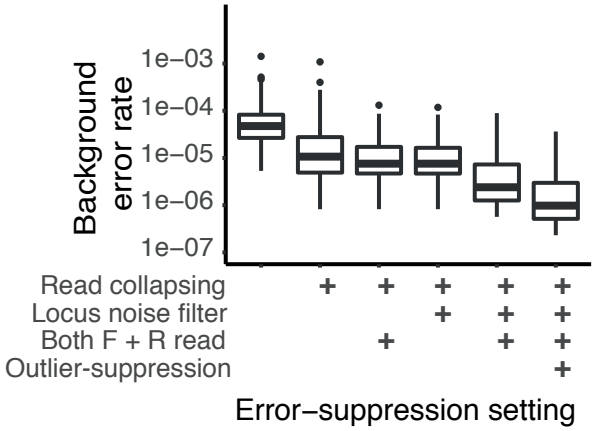


Fig. 4.5 **Reduction of error rates following error-suppression methods**
Each mutation was required to be observed in both mates of a paired-end sequencing read, and the given locus must be characterised as low-noise in control samples (Method 4.7.3). Error rates were suppressed to 1 mismatch per 10^7 bases.

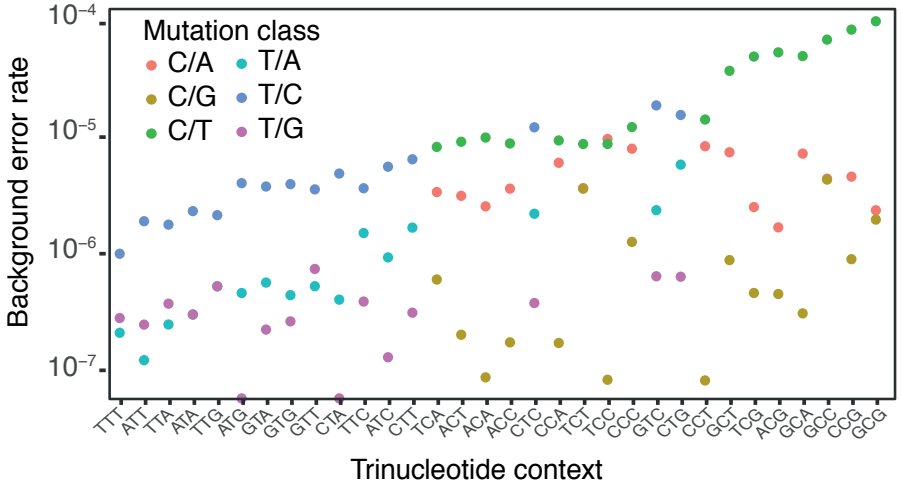


Fig. 4.6 **Error rates by trinucleotide context and mutation class following data filtering**
Mismatch rates per trinucleotide context were determined (Method 4.7.4). Error rates varied by more than an order of magnitude within the same mutation class, highlighting the importance to assess loci with respect to their trinucleotide context.

Signal-enrichment

Previous studies have shown relationships between tumour allele fraction and plasma allele fraction [124, 125], and have shown size differences between mutant and wild-type cfDNA fragments [65, 198]. To effectively use sequencing information, INVAR enriches for ctDNA signal through probability weighting based on ctDNA fragment sizes (Figs. 4.7) and each mutation's allele fraction in tumour sequencing data (Fig. 4.8, Method 4.7.6). After weighting, a significance level is generated for each of the loci in the patient-specific mutation list, which are combined into an aggregate likelihood function. The likelihood is converted to the INVAR score through comparison to the cumulative distribution of likelihoods for control samples, such that the INVAR score represents the estimated specificity level (1 minus the false positive rate) at which a sample would be 'detected' with this score (Method 4.7.9). An integrated mutant allele fraction (IMAF) is determined using a maximum likelihood estimator (Method 4.7.8).

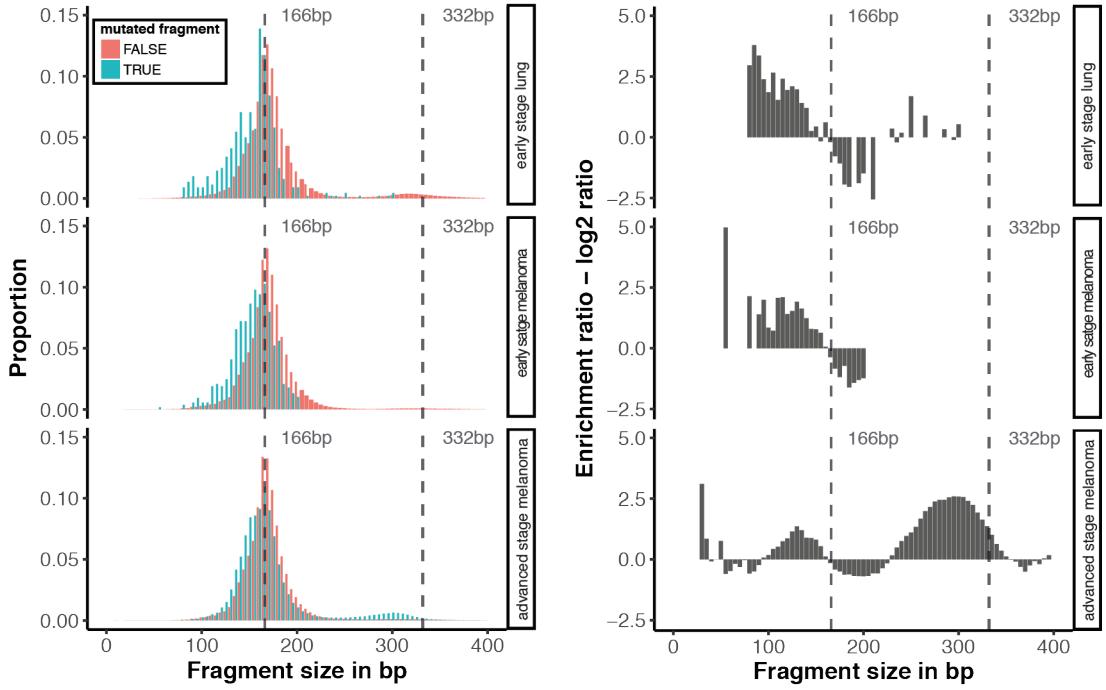


Fig. 4.7 **Fragment size profiles and enrichment ratios**
Left panel, for each cohort, size profiles were generated for mutant and wildtype fragments. The proportion of data per bin is shown. Wild-type data are shown in red, and mutant data are shown in blue. The size of a nucleosome (166bp) is indicated with a dashed line, as is the di-nucleosome (332bp).
Right panel, for each size range, the ratio between mutant and wild-type fragments for each of the three cohorts was determined. This ratio between mutant and wild-type fragments for a given size range indicates the extent of mutant fragment enrichment per bin. Size ranges enriched for ctDNA are assigned more weight by the INVAR algorithm (Method 4.7.6).

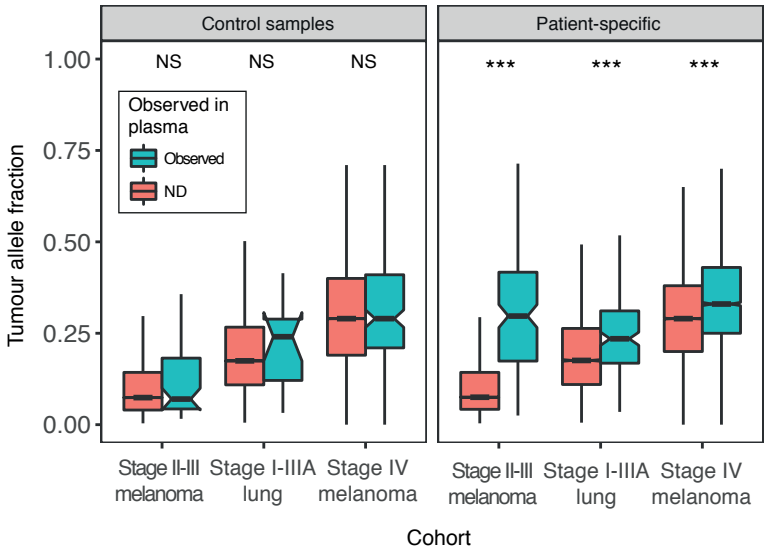


Fig. 4.8 **Tumour allele fractions of loci observed in plasma**
Tumour allele fractions were compared between loci with vs. without detected signal in plasma, showing that loci with signal in plasma have a significantly higher tumour allele fraction in patient samples. There was no significant increase in tumour allele fraction when the same analysis was performed on non-patient-specific samples, i.e. in patient samples but at loci not belonging to that patient. Student's t-test was performed. NS, not significant; ***, $p < 0.0001$.

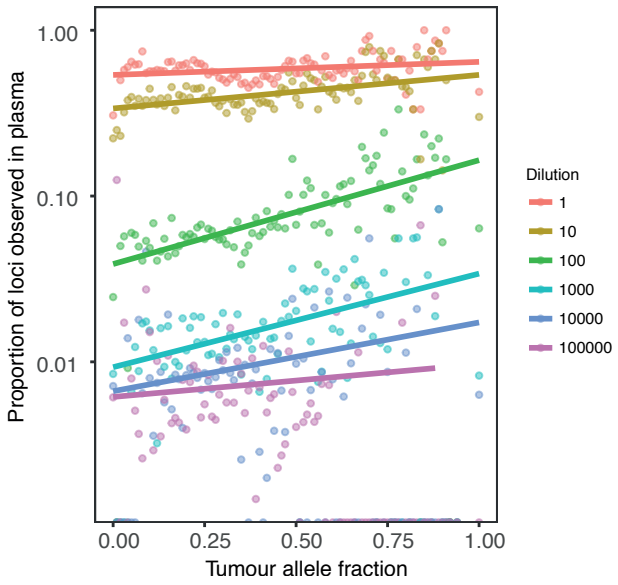


Fig. 4.9 **Detection rate of mutations by tumour allele fraction**
Using error-suppressed data, tumour loci were grouped into bins of 0.01 mutant allele fraction, and the proportion of loci observed in plasma was determined for different levels of a dilution series. The dilution level of the spike-in dilution series is indicated by each colour. At each dilution level, there is a positive correlation between the tumour allele fraction and proportion of loci observed in plasma.

4.4.2 Analytical performance of the INVAR algorithm

Dilution series

We evaluated the analytical performance of INVAR by analysis of sequencing from a custom capture panel, in a dilution series of plasma from one melanoma patient, for whom we had identified 5,073 mutations through exome sequencing (Method 4.6.9), diluted into plasma from healthy control volunteers. Without error-suppression, the lowest dilution detected with INVAR score >0.92 (equivalent to a specificity, calculated in Method 4.7.6) had an IMAF of 10^{-4}. Following error-suppression and size-weighting, samples diluted in healthy control cfDNA down to an expected IMAF of 1 x 10^{-6} (1 part per million, ppm) were detected with an INVAR score >0.92. Of 3 replicates diluted to 1 ppm, two were detected with IMAF values of 9.2 ppm and 0.47 ppm. The

correlation between IMAF and the expected mutant fraction was of 0.99 (Pearson's r, $P = 4.2 \times 10^{-12}$, Fig. 4.10).

We downsampled the sequencing data *in silico* to include only subsets of the patient-specific mutation list (Method 4.6.9), which confirmed that a larger number of mutations and higher IR resulted in higher ctDNA detection rates (Fig. 4.11).

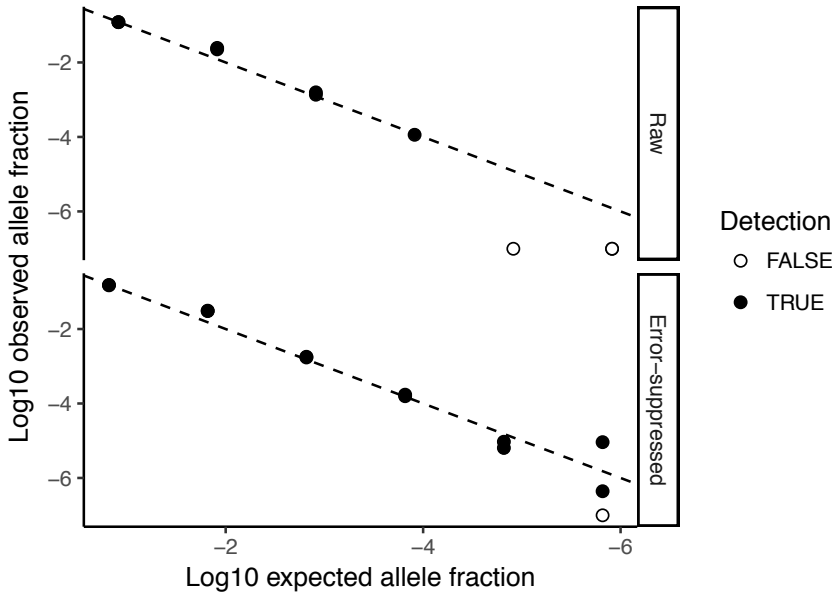


Fig. 4.10 **Spike-in dilution experiment of 5,073 patient-specific mutations to test the sensitivity of INVAR**
In non-error-suppressed data with INVAR, samples with allele fractions at or above 10^{-4} were detected. Using error-suppressed data with INVAR, ctDNA was detected in replicates for all dilutions above 10^{-6}, and in 2 of 3 replicates at the expected ctDNA concentration of 1×10^{-6}. The dashed line indicates the allele fractions expected for every sample based on the dilution factor.

Application on exome and WGS data

To test the ability to detect low levels of ctDNA with other types of sequencing data, we also applied INVAR to WES data (median depth 238x) and sWGS data (median depth 0.6x) generated from samples with IMAF down to 10^{-4}. In exome sequencing

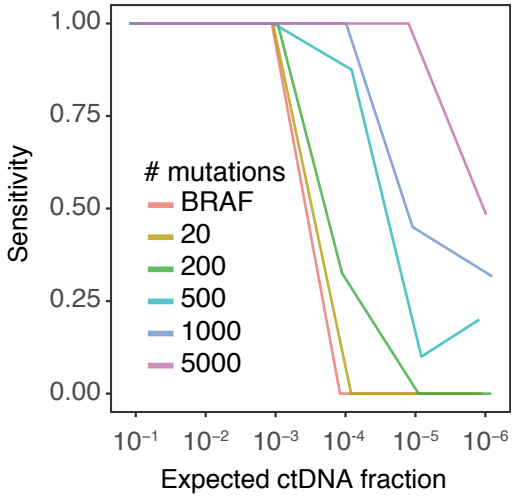


Fig. 4.11 **Sensitivity analysis with varying numbers of mutations**
The number of loci analysed was downsampled *in silico* to between 1 to 5,000 mutations with the highest observed mutant allele fraction in tissue (Method 4.6.9). With 5,000 mutations, ctDNA was detected at 10^{-6} with 50% sensitivity.

data, ctDNA was detected in all samples (down to IMAF of 10^{-4}) and the IMAF values between custom capture and exome sequencing showed a correlation of 0.96 (Pearson's r, P = 9 x 10^{-12}, Fig. 4.12). When WGS data was used, samples with ctDNA mutant allele fractions down to 10^{-4} measured by custom capture were detected, although they were observed at higher allele fractions due to the limited number of IR achieved using sWGS.

These data show that the INVAR algorithm can be applied flexibly to data that is not custom capture. Any sequencing data with reads overlying patient-specific sites may be used, either by capture based or untargeted methods. By using 0.6x WGS for this analysis, many of the loci have zero reads due to the low depth, but signal may still be detected at the loci with signal.

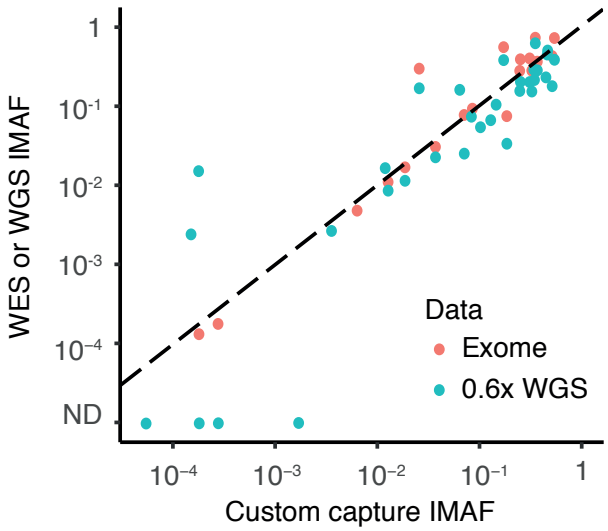


Fig. 4.12 **Generalisation of INVAR to exome and sWGS**
In addition to designing custom sequencing panels, we tested the application of exome and low-coverage whole genome sequencing to a subset of samples with IMAF down to 10^{-4}, achieving detection in 100% of samples undergoing whole exome sequencing. Correlation in IMAF between exome and custom capture was 0.96 (Pearson's r, P = 1.3 x 10^{-11}). Using low-coverage whole genome sequencing, we detected and quantified ctDNA down to IMAFs below 10^{-2}.

4.4.3 Quantification to parts per million in clinical samples

We applied INVAR to sequencing data generated using custom capture panels from 160 plasma samples from 47 stage II-IV melanoma patients and 19 plasma samples from 19 stage I-IIIA NSCLC patients. We analysed a median of 625 mutations per patient with melanoma and 388 per patient with stage I-IIIA NSCLC, resulting in up to 2.9 x 10^6 IR per sample (median 1.7 x 10^5 IR), thus analysing orders of magnitude more cell-free DNA fragments compared to methods that analyse individual or few loci. Each sample analysed in this section is plotted in 2-dimensional space based on the number of mutations and haploid genomes analysed in Supplementary Fig. 4.36. We detected ctDNA levels ranging from 0.98 x 10^{-6} to 0.57 (Fig. 4.13, Tables 4.4, 4.5), confirming a dynamic range of nearly 6 orders of magnitude, and detection of trace levels of ctDNA in plasma samples from cancer patients to the ppm range, from a median of 1638 copies of the genome (5.46 ng of DNA; Table 4.4). ctDNA was thus confidently detected by INVAR in samples containing as little as 49 femtograms of tumour DNA (0.007 of the amount of DNA in a single tumour cell). Given the limited input, these ctDNA levels would be below the 95% limit of detection for a perfect single-locus assay in 34% of the cases, which is indicated by the presence of a filled circle in Fig. 4.13.

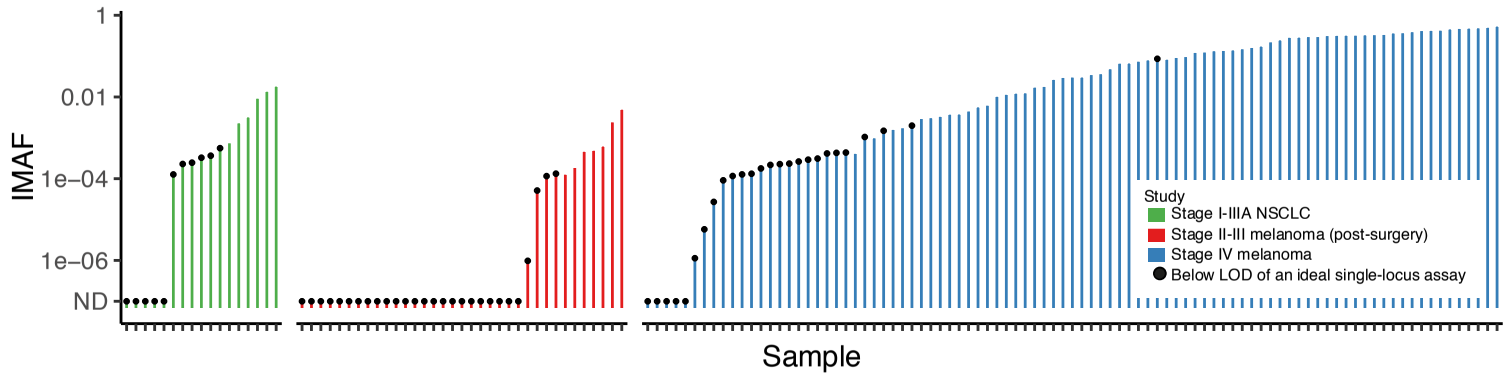


Fig. 4.13 **Distribution of IMAF values in this study**
IMAFs are shown in ascending order for each sample in this study. Given the limited input, ctDNA levels would be below the 95% limit of detection for a perfect single-locus assay in 39% of the cases, indicated by filled circles. Bars are coloured by cohort. ND, sample not detected.

4.4.4 Comparison against clinical markers in advanced melanoma

Next, we compared the IMAF assessed by INVAR to other metrics for ctDNA levels and other clinical markers in patients with unresectable stage IV melanoma. We compared IMAF to the variant allele fraction (VAF) for ctDNA that was determined by amplicon sequencing of an individual mutation, showing a Pearson's correlation of 0.85, (Fig. 4.14). Amplicon sequencing of *BRAF* or *NRAS* alone detected ctDNA in 73% of longitudinal samples, with a minimum detected VAF of 0.01%, whereas INVAR detected ctDNA in 95.3% of longitudinal samples at INVAR score of 1.00 (Supplementary Fig. 4.38).

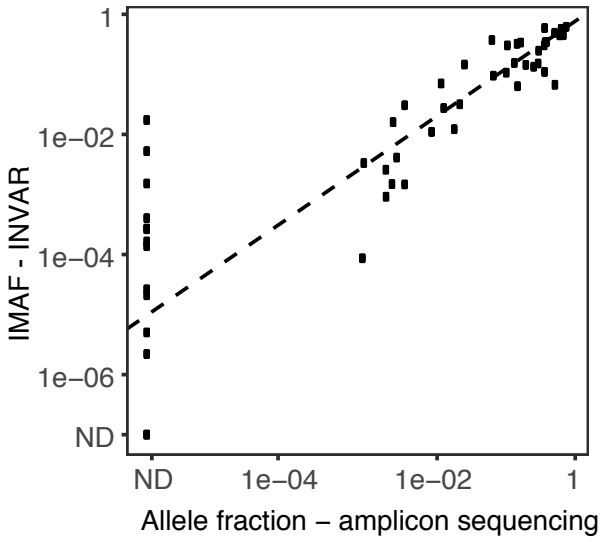


Fig. 4.14 **Comparison of INVAR IMAF values against amplicon sequencing**
Amplicon sequencing was performed on a subset of samples with TAm-Seq [29], targeting one mutation per patient, either a *BRAF* or *NRAS* mutation. A Pearson correlation of 0.85 was observed between the two methods.

IMAF was compared to tumour volume from CT imaging (Table 4.6,Method 4.6.10) and to serum lactate dehydrogenase (LDH), a routinely used clinical biomarker for melanoma. For one patient treated with vemurafenib, ctDNA was detected at IMAF as low as 1.1 ppm, and the corresponding tumour volume was 1.3 cm^3 (Fig. 4.15). We found a correlation between IMAF and imaging of 0.8 (Pearson's r, P = 6.7 x 10^{-10},

Fig. 4.16), and a correlation of 0.82 between IMAF and LDH (Pearson's r, P = 4 x 10^{-13}, Fig. 4.17). Compared to other studies [111, 117], INVAR showed a steeper gradient between tumour volume and IMAF, which may reflect the lower detectable IMAF with INVAR.

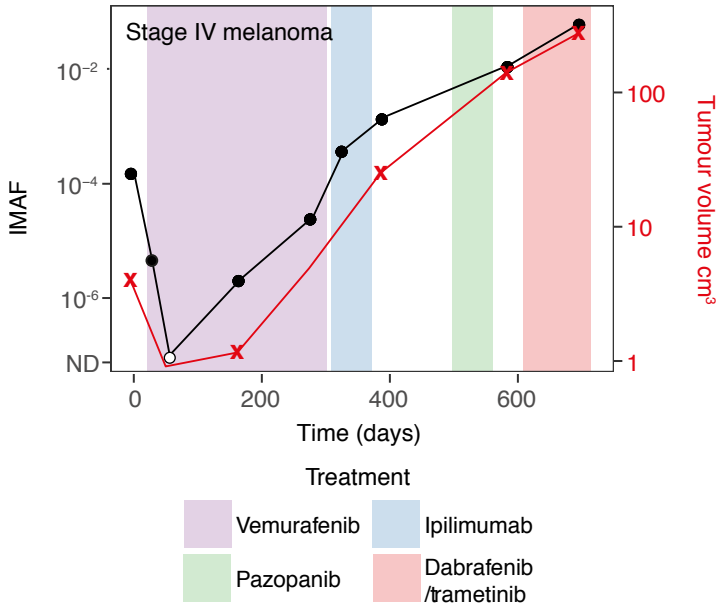


Fig. 4.15 **INVAR monitors disease to parts per million and 1.3cm^3**
The shaded red box indicates the working point of INVAR using millilitres of plasma. ctDNA was detected to 1.1 ppm during treatment with anti-*BRAF* targeted therapy, when disease volume was on the order of 1cm^3. ctDNA IMAF and tumour volume in cm^3 are plotted over time. Treatments are indicated by shaded boxes. ND, not detected.

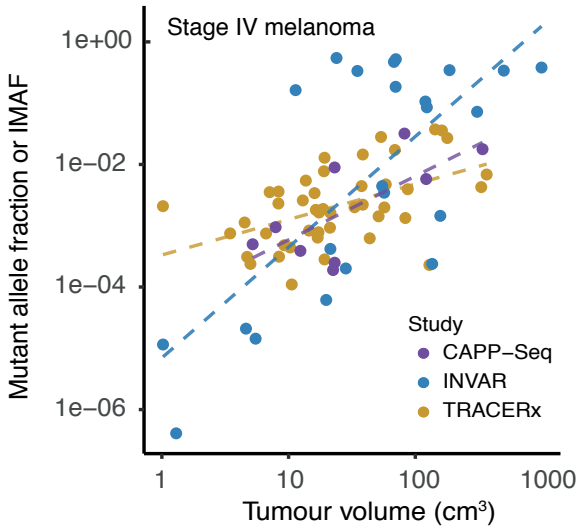


Fig. 4.16 **Comparison between ctDNA and tumour volume in this study and others**
ctDNA IMAF values were compared against tumour volume measurements by CT. These data were also compared against previous publications measuring multiple mutations per patient in NSCLC, using CAPP-Seq [111], and using multiplexed PCR in the TRACERx cohort [117].

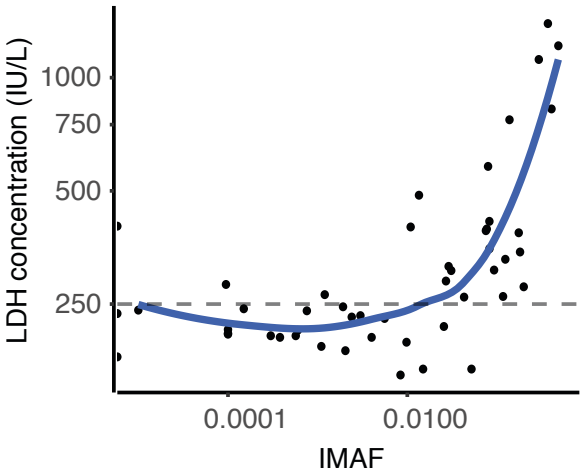


Fig. 4.17 **Correlation between serum lactate dehydrogenase with IMAF in advanced stage melanoma patients**
A Pearson correlation score of 0.46 was observed between serum LDH and IMAF (P = 0.0058). A dashed line is drawn at 250 IU/L, the upper limit of normal for LDH. A smoothed line was fitted to the data.

4.4.5 Detection of stage I-IIIA NSCLC

We tested ctDNA detection by INVAR in plasma samples collected pre-treatment from 19 patients with newly diagnosed stage I-IIIA NSCLC (consisting of 11, 6 and 2 patients with stage I/II/IIIA). In two samples, ctDNA was not detected, but fewer than 2 x 10^4 IR were analysed due to a small number of mutations identified in WES of matched tissue (59 and 93 in each case) and these were considered as non-evaluable and discarded due to limited effective sensitivity (Method 4.7.9). Of the 17 evaluable patients, the median number of IR was 5.6 x 10^4 (IQR 3.5-9.6 x 10^4). ctDNA was detected in 12 out of 17 patients, including 2/5 patients with stage IA, 3/5 patients with stage IB, 5/5 patients with stage II and 2/2 patients with stages III (Table 4.7). We performed receiver operating characteristic (ROC) analysis to test different thresholds of INVAR score to define detection in these patient and control samples (including both plasma samples from healthy volunteers, and data from patients across non-matched mutation lists, Method 4.7.9). The area under the curve (AUC) values were 0.72, 0.83 and 0.995 for stage I only, stage I-IIIA, and stages II-IIIA only, respectively (Fig. 4.18).

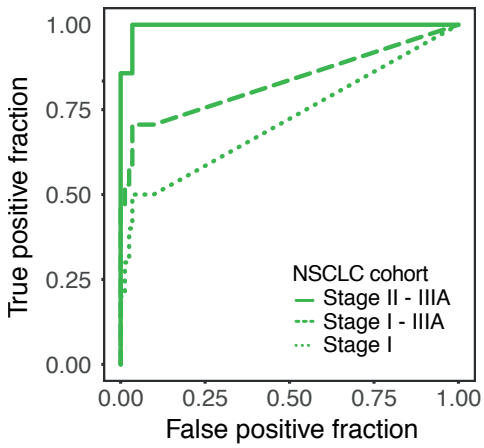


Fig. 4.18 **ROC analysis of stage I-IIIA NSCLC cohort**
ROC analysis was perfomed on the INVAR scores generated from stage I-IIIA NSCLC patient samples. AUC values of 0.72, 0.83 and 0.995 were achieved for stage I (dotted line), stage I-IIIA (dashed line) and stages II-IIIA (solid line), respectively. The ground truth was defined by analysing a set of healthy individuals alongside the NSCLC patient samples.

4.4.6 Detection of ctDNA post-surgery in stage II-III melanoma

To test INVAR in the residual disease setting, we analysed samples from 38 patients with resected stage II-III melanoma recruited in the UK AVAST-M trial (ISRCTN81261306, Table 4.8), collected up to 6 months after surgery with curative intent. We interrogated a median of 1.65 x 10^5 IR (range 0.62 x 10^5 to 3.23 x 10^5) and detected ctDNA (with INVAR score >0.93) to a minimum IMAF of 0.98 ppm. Of the evaluable patients (n = 33, Method 4.7.9), ctDNA was detected in 50% of stage II-III patients who later recurred, and was associated with a significantly shorter disease-free interval (4.5 months vs. median not reached with 5 years' follow-up; Hazard ratio (HR) = 3.69; 95% CI 1.44-9.46, P = 0.007; Fig. 4.19) and overall survival (2.6 years vs. median not reached, Fig. 4.20). Using this method, the sensitivity of ctDNA to predict recurrence was 50%, in contrast to previous analysis of plasma from 161 patients from the same trial with resected *BRAF* or *NRAS*-mutant melanoma, where ctDNA was detected in 15.6% of patients who later relapsed using a single-locus digital PCR assay targeting the respective *BRAF*/*NRAS* mutations [266].

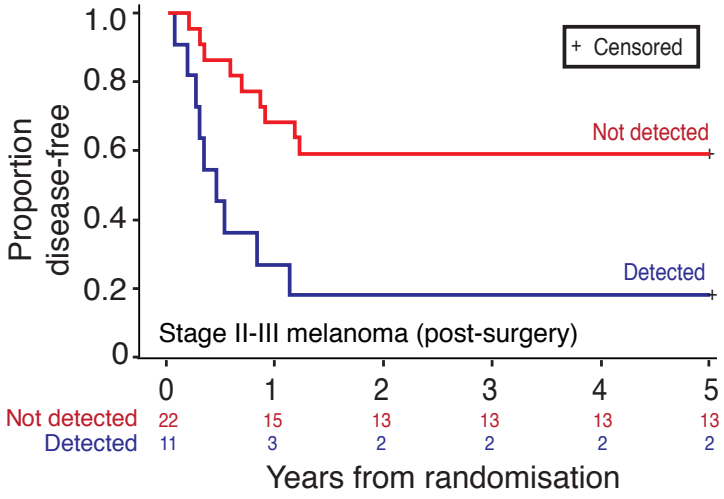


Fig. 4.19 **Kaplan-Meier analysis of stage II–III melanoma patients for disease-free interval**
Kaplan-Meier analysis was performed based on ctDNA detection using INVAR. Disease-free interval was significantly poorer in patients with ctDNA detected within 6 months following surgery (P = 0.007, Cox proportional hazards model).

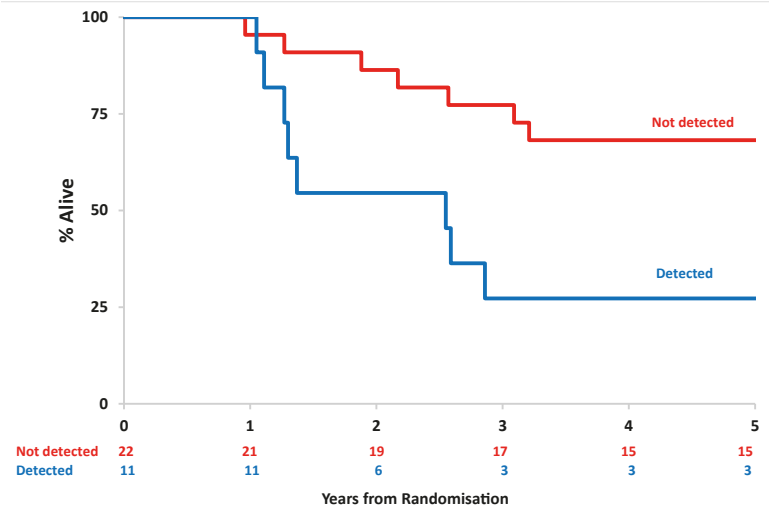


Fig. 4.20 **Kaplan-Meier analysis of stage II–III melanoma patients for overall survival**
Patients with detected ctDNA had a significantly poorer overall survival (P = 0.02, Cox proportional hazards model). The median survival for patients with detected ctDNA was 2.6 years (95% CI 1.1-5.3) vs. median not reached (95% CI 3.1 – median not reached).

4.4.7 Estimation of detection rates with varying IR

Using IMAF values from the clinical samples, we estimated the expected detection rates for different cohorts of patients with fewer IR, and fitted a linear model (R^2 = 0.95). In stage IV melanoma patients prior to treatment, we detected ctDNA in 100% of cases with 10^5 IR (Fig. 4.21). In patients with stage IV melanoma undergoing treatment, where ctDNA levels are lower, an extrapolation of the linear fit predicted 10^6 IR would enable detection of ctDNA in nearly all samples (Supplementary Fig. 4.39). In patients with early-stage NSCLC, a linear extrapolation predicted that it may be possible to detect ctDNA in nearly all patients if ~10^6 IR were sequenced for each sample. In contrast, for stage II-III melanoma patients who undergo surgery, even analysis of 10^7 IR is predicted to result in detection of ctDNA within 6 months of surgery in only ~60% of patients who would relapse. Such depth of analysis may be currently impractical based on sequencing costs and the amount of input DNA required (Fig. 4.21).

4.4.8 Application of INVAR in limited input samples

Sensitive monitoring using sWGS data

We explored the applicability of INVAR to other scenarios, including low-input or low-depth sequencing from plasma samples of cancer patients. High-sensitivity detection of ctDNA from such sequencing data could potentially enable a generic and widely applicable monitoring assay, and could also be applied to analyse minuscule blood volumes, such as can be obtained from a pin-prick or dried blood spot (Fig. 4.22). We showed that libraries can be generated from individual genome copies (Supplementary Fig. 4.40) to demonstrate the feasibility of generating data from blood spots. Next, we applied INVAR to sWGS data (0.6x coverage) from longitudinal samples from

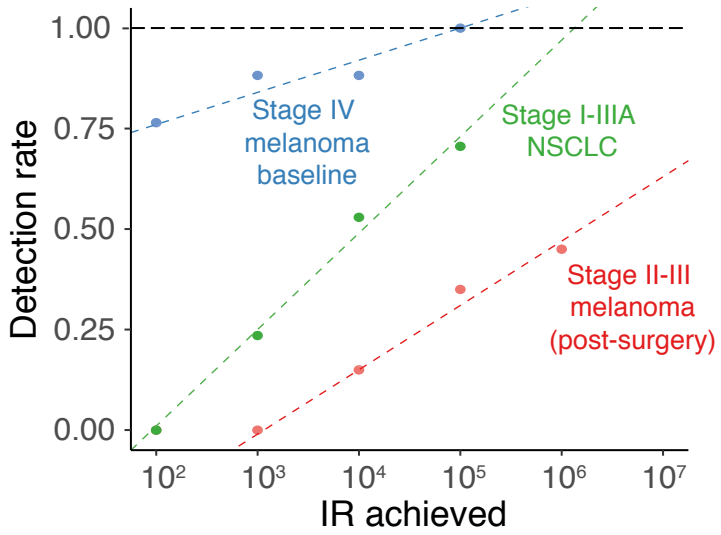


Fig. 4.21 **Estimated the detection rates of ctDNA for different IR sequenced**
We observe a linear relationship (R2 = 0.95) between the number of IR and detection rate in the baseline samples of the stage IV melanoma cohort (blue). ctDNA was detected in 100% of these samples with 10^5 IR. Detection of ctDNA in follow-up samples required a larger number of IR (Supplementary Fig. 4.39). In stage I-IIIA NSCLC at diagnosis (green) and stage II-III melanoma post-surgery (red), linear relationships were observed between IR and detection rates, and the predicted rates of detection of ctDNA were extrapolated.

patients with stage IV melanoma, where ctDNA was quantified down to an IMAF of 10^{-4} (Fig. 4.23). Analysis of 0.6 genomes (0.6 hGA) could hypothetically be obtained from a sequencing library prepared from ~1-2 genome copies (after losses from library preparation and sequencing), equivalent to the DNA from ~2µL plasma with a concentration of cfDNA of between 1-5ng/mL [59].

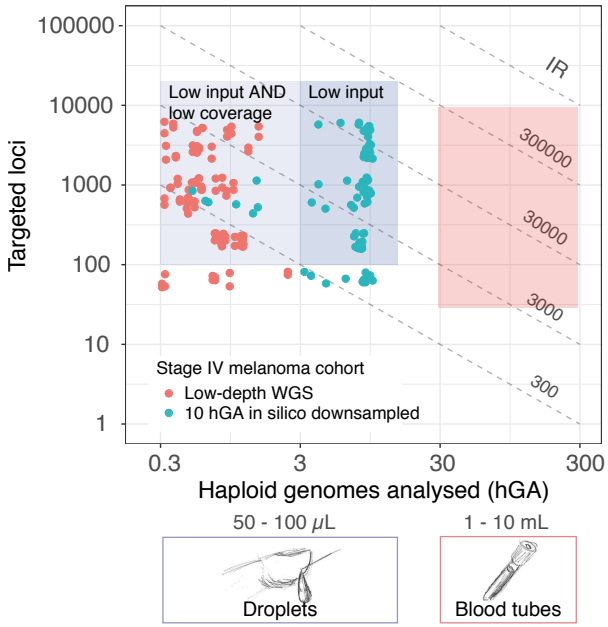


Fig. 4.22 **Number of hGA vs. mutations targeted - low IR setting**
Samples were either downsampled *in silico* to 10 hGA, or libraries were sequenced with 0.6x WGS, equivalent to 0.6 hGA. 10 hGA were chosen to simulate the deduplicated sequencing coverage obtainable from a single droplet of blood or blood spot (volume = 75μL, plasma content = 55%, cfDNA concentration = 1-5ng/mL, library preparation efficiency = 30%, giving 4-20 hGA per droplet/blood spot). The dark blue shaded box indicates the working point achieved by INVAR with low input. The light blue shaded box indicates the potential for monitoring using sWGS.

Detection and monitoring from 10 genome copies *in silico*

Theoretically, one blood drop may yield up to 60 genome copies, assuming a droplet volume of 75µL, plasma content of 55%, cfDNA concentration of 1-5ng/mL [279]; with a library preparation efficiency of 30%, this may yield up to 4x-20x deduplicated

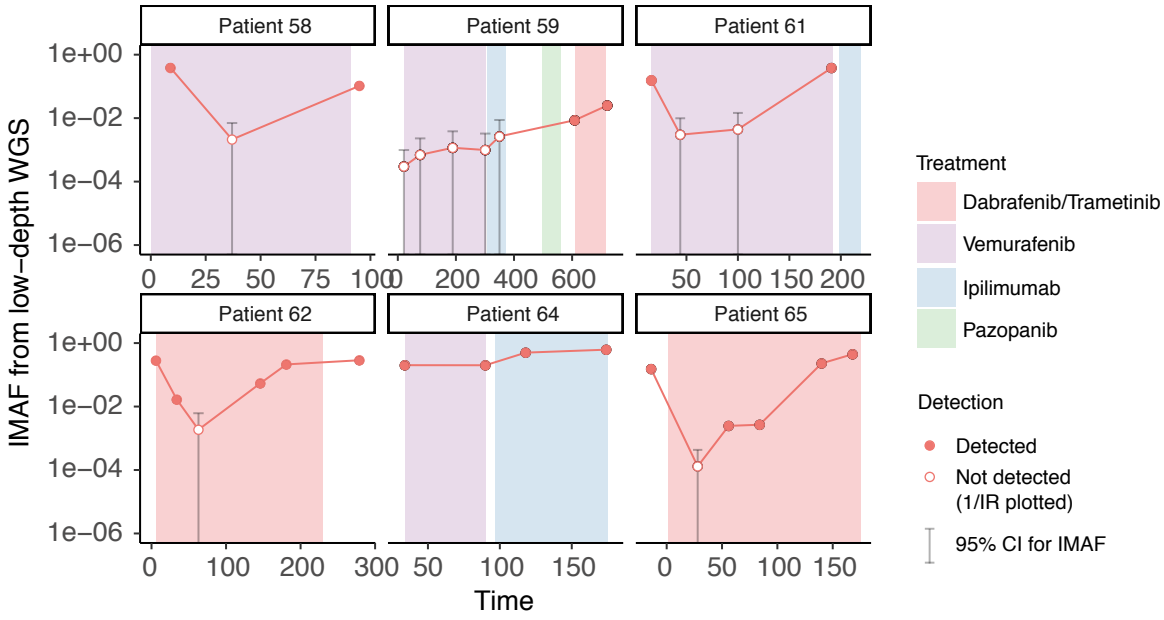


Fig. 4.23 **Longitudinal monitoring of ctDNA by applying INVAR to sWGS**
For 6 patients with >500 mutations, sWGS data was generated and deduplicated. The solid lines and circles indicate low-depth IMAF values for each time point from sWGS. Filled circles indicate detection at INVAR score >0.99. For samples with lower INVAR score, the 95% confidence intervals of the ctDNA level are shown based on the number of IR interrogated. ND, not detected, indicating the maximal likely IMAF with 95% confidence intervals based on the number of IR sequenced.

coverage. We therefore downsampled plasma sequencing data from stage IV melanoma patients *in silico* to 10 hGA, to simulate the analysis of cfDNA from a droplet of blood. We detected ctDNA in 90% of cases to concentrations of 10^{-4} (Fig. 4.24), and used this data to monitor ctDNA levels in longitudinal plasma samples from these patients (Supplementary Fig. 4.41).

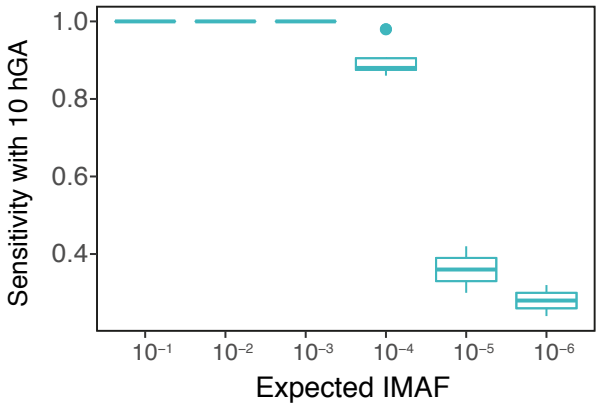


Fig. 4.24 **Assessment of sensitivity with 10 hGA with a spike-in dilution series**
For 6 patients with >500 mutations, deduplicated sWGS data were generated. The same spike-in dilution series (Method 4.6.9) was downsampled to 10 hGA, and sensitivity for detecting ctDNA at estimated specificity (INVAR score) of 0.97 using INVAR was determined.

Different cancer types have different rates of mutations across the genome [234]. These would result in varying numbers of IR in data from plasma cfDNA sequencing. With WGS data available for each patient's tumour and WGS data from 10 hGA of cell-free DNA (obtainable from two blood droplets or spots), INVAR could theoretically detect ctDNA at levels ranging from 4 to 330 ppm (Fig. 4.25). Previous work has shown the feasibility of detecting fetal DNA in cell-free DNA samples collected prenatally on dried blood spots [280, 281]. Given the sensitivity of detection we have shown using INVAR analysis, it may in future be possible to monitor cancer by analysis of DNA collected on dried blood spots to detect low concentrations of ctDNA, potentially enabling self-collection of blood samples by patients with appropriate devices.

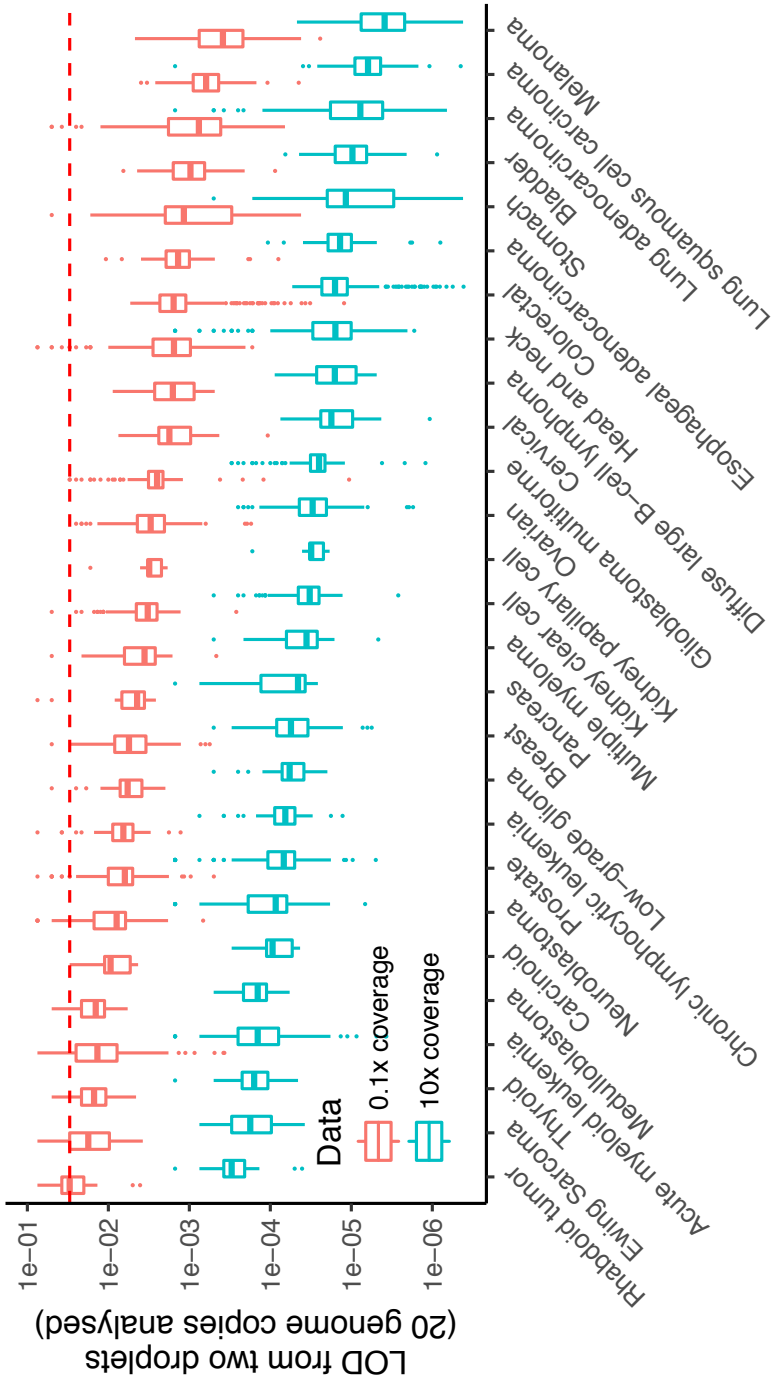


Fig. 4.25 **Potential limits of detection for ctDNA from two droplets of blood in different cancer types**
Based on known mutation rates per Mbp of the genome for different cancer types [234], the number of IR per droplet can be estimated. If low-depth coverage is used, with sufficient numbers of mutations in the tumour WGS, ctDNA can be detected to fractional concentrations of 10^{-4} at high specificity and sensitivity using INVAR analysis of single nucleotide variants, in contrast to methods based on copy number alterations which have detected ctDNA down to 3% [110, 115].

4.5 Discussion

In summary, we described an algorithm that can achieve high levels of sensitivity by integration of variant reads across hundreds to thousands of mutated loci. We demonstrated ctDNA detection to parts per million, 1-2 orders of magnitude lower than previous data [117, 170, 203, 277, 282]. In addition, this method enables the detection of low concentrations of ctDNA in limited input samples, potentially even from a few drops of blood. We showed that the INVAR algorithm can be applied to plasma sequencing data from custom capture panels, exome sequencing and WGS. Leveraging information from tumour sequencing, INVAR enables monitoring of ctDNA levels from sWGS data with greater sensitivity than previously-described methods based on copy-number analysis [114, 115]. As tumour sequencing becomes increasingly implemented in personalised oncology, these data may be leveraged with INVAR to enhance ctDNA monitoring, through limited plasma volumes and to greater levels of sensitivity.

4.6 Experimental Methods

4.6.1 Patient cohort

Samples were collected from patients enrolled on the MelResist (REC 11/NE/0312), AVAST-M (ISRCTN81261306)30 and LUCID (REC 14/WM/1072) studies. MelResist is a translational study of response and resistance mechanisms to systemic therapies of melanoma, including *BRAF* targeted therapy and immunotherapy, in patients with stage IV melanoma. AVAST-M is a randomised control trial which assessed the efficacy of bevacizumab in patients with stage IIB-III melanoma at risk of relapse following surgery; only patients from the observation arm were selected for this analysis. LUCID is a prospective and observational study of stage I-IIIB non-small cell lung cancer patients (NSCLC) who are planning to undergo radical treatment (surgery or radiotherapy +/- chemotherapy) with curative intent. All studies were coordinated by the Cambridge Cancer Trials Unit-Cancer Theme, and demographics and clinical outcomes were collected prospectively.

4.6.2 Sample collection and processing

For patients with stage I-III melanoma or NSCLC, formalin-fixed, paraffin-embedded tumour blocks and matched buffy coat samples were obtained. Fresh frozen tumour biopsies prior to treatment were collected from patients with stage IV cutaneous melanoma. For patients on the AVAST-M study, plasma samples were collected within 12 weeks of tumour resection, with a subsequent sample after 3 months, where available. Patients on the LUCID study had one plasma and matched buffy coat sample taken pre-surgery. Longitudinal samples were collected during treatment of patients with stage IV melanoma as part of the MelResist study. Peripheral blood samples were

collected at each clinic visit in S-Monovette 9mL EDTA tubes. For plasma collection, samples were centrifuged at 1600 g for 10 minutes within an hour of the blood draw, and then an additional centrifugation of 20,000 g for 10 minutes was carried out. All aliquots were stored at -80°C.

4.6.3 Tissue and plasma extraction

FFPE samples were sectioned into up to 8 μm sections, and one H&E stained slide was generated, which was outlined for tumour regions by a histopathologist. Marked tumour regions were macrodissected, and DNA extraction was performed using the QIAamp DNA FFPE Tissue Kit using the standard protocol, except with incubation at 56°C overnight and 500 rpm agitation on a heat block. DNA was eluted twice using 20 μL ATE buffer each time with centrifugation at full speed. Following extraction, DNA repair was performed using the NEBNext FFPE DNA Repair Mix as per the manufacturer's protocol.

Fresh frozen tissue biopsies were first homogenised prior to DNA extraction, which was performed as follows: up to 30 mg of each fresh frozen tissue biopsy sample was combined with 600 μL RLT buffer, then placed in a Precellys CD14 tube (Bertin Technologies) and homogenised at 6,500 rpm for two bursts of 20 seconds separated by 5 seconds. Subsequently, the Qiagen AllPrep extraction kit as per the manufacturer's protocol.

Genomic DNA was extracted from up to 1 mL whole blood or buffy coat using the Gentra Puregene Blood Kit (Qiagen) as per the manufacturer's protocol. Samples were eluted in two rounds of 70 μL buffer AE and incubated for 3 minutes before centrifugation. Up to 4mL of plasma was extracted using the QIAsymphony (Qiagen) with a QIAamp protocol. DNA was eluted in 90 μl elution buffer and stored at -80°C.

Plasma samples were extracted using the QIAsymphony instrument (Qiagen) using a 2-4mL QIAamp protocol. For each QIAsymphony batch, 24 samples were extracted, which included a healthy individual control sample (Seralab).

4.6.4 DNA quantification

Following extraction of fresh frozen, FFPE and genomic DNA eluted DNA concentration was quantified using a Qubit fluorimeter with a dsDNA broad range assay (ThermoFisher Scientific). To quantify cfDNA concentration of plasma DNA eluates, digital PCR was carried out using a Biomark HD (Fluidigm) using a Taq-man probe for the housekeeping gene RPP30 (Sigma Aldrich). 55 PCR cycles were used. The RPP30 assay was 65 bp in length. The estimated number of RPP30 DNA copies per μl of eluate was used to determine the cfDNA concentration in the original sample.

4.6.5 Library preparation

Tumour

FFPE tumour tissue DNA samples (up to 150 ng) and buffy coat DNA samples (75 ng) were sheared to a length of 150, using the Covaris LE 220 (Covaris, Massachusetts, USA). The standard Covaris protocol for a final fragment length of 150bp and an input volume of 15μl using the 8 microTUBE-15 AFA Beads Strip V2 was used. After the shearing, the fragmentation pattern was verified using a Bioanalyser (Agilent). Sequencing libraries were prepared using the ThruPLEX DNA-seq kit (Rubicon). 100ng and 50ng sheared tumour and buffy coat DNA were used, respectively, and the protocol was carried out according to the manufacturer's instructions. The number of amplification cycles was varied during library preparation according to the manufacturer's recommendations. Fresh frozen tumour biopsies and matched buffy coat library

preparation was performed as described by Varela et al. [274] using the SureSelectXT Human All Exon 50 Mb (Agilent) bait set. Samples were multiplexed and sequenced with a HiSeq 2000 (Illumina).

Library concentration was determined using qPCR with the Illumina/ROX low Library Quantification kit (Roche). Library fragment sizes were determined using a Bioanalyser (Agilent). After library preparation, exome capture was performed with The TruSeq Exome Library Kit (Illumina), using a 45Mbp exome baitset. Three libraries were multiplexed in one capture reaction and 250ng of each library was used as input. For compatibility with ThruPLEX libraries, the protocol was altered by adding 1μl of i5 and i7 TruSeq HT xGen universal blocking oligos (IDT) during each hybridisation step. To compensate for the increased hybridisation volume, the volume of CT3 buffer was adjusted to 51 μl. Two rounds of hybridisations were carried out, each lasting for 24 hours. Library QC was performed using qPCR and Bioanalyser, as above. Samples were multiplexed and sequenced with a HiSeq 4000 (Illumina).

Plasma

cfDNA samples were vacuum concentrated at 30°C using a SpeedVac (ThemoFisher) prior to library preparation where required. Library preparation for plasma cfDNA was performed using the Rubicon ThruPLEX Tag-Seq kit. The number of PCR amplification cycles during the ThruPLEX protocol was varied between 7-15 cycles, as recommended by the manufacturer. Following amplification and sample barcoding, libraries were purified using Ampure XT beads (Beckman Coulter) at a 1:1 ratio. Library concentration was determined using the Illumina/ROX low Library Quantification kit (Roche). Library fragment sizes were determined using a Bioanalyser (Agilent).

For the stage IV melanoma cohort, library preparation and sequencing was run in duplicate to assess the technical reproducibility of the experimental and computational

method, showing a correlation between IMAF values generated by the INVAR pipeline of 0.97 (Pearson's r, $P < 2.2 \times 10^{-16}$). For the early-stage cohorts, input cfDNA material was not split and was instead prepared and sequenced as a single sample per time point.

4.6.6 Custom hybrid-capture panel design and sequencing of plasma

Following tumour mutation calling (Method 4.7.1), custom hybrid-capture sequencing panels were designed using Agilent SureDesign software. Between 5 and 20 patients were grouped together per panel in this implementation. Baits were designed with 4-5x density and balanced boosting for melanoma patients and 1x density and balanced boosting for lung cancer patients. 95.5% of the variants had baits successfully designed; bait design was not reattempted for loci that had failed. Custom panels ranged in size between 1.26-2.14 Mb with 120 bp RNA baits, designed based on patient-specific mutation lists generated by mutation calling (Fig. 4.26, Method 4.7.1). Individualised capture was performed on 213 libraries (Table 4.4).

Libraries were captured either in single or 3-plex (to a total of 1000 ng capture input) using the Agilent SureSelectXT protocol, with the addition of i5 and i7 blocking oligos (IDT) as recommended by the manufacturer for compatibility with ThruPLEX libraries [283]. Custom Agilent SureSelectXT baits and 13 cycles of post-capture amplification were used. Post-capture libraries were purified with Ampure XT beads at a 1:1.8 ratio, then were quantified and library fragment size was determined using a Bioanalyser (Agilent).

4.6.7 Exome capture sequencing of plasma

For exome sequencing of plasma, the Illumina TruSeq Exome capture protocol was followed. Libraries generated using the Rubicon ThruPLEX protocol (as above) were

pooled in 3-plex, with 250ng input for each library. Libraries underwent two rounds of hybridisation and capture in accordance with the protocol, with the addition of i5 and i7 blocking oigos (IDT) as recommended by the manufacturer for compatibility with ThruPLEX libraries [283]. Following target enrichment, products were amplified with 8 rounds of PCR and purified using AMPure beads prior to QC.

4.6.8 Low-depth whole-genome sequencing of plasma

For WGS, 30 libraries were sequenced per lane of HiSeq 4000, achieving a median of 0.6x deduplicated coverage per sample. For these libraries, since the number of informative reads (IR) would limit sensitivity before background errors would become limiting, we used error-suppression with family size 1 for this particular setting. Error rates per trinucleotide were compared between WGS and custom hybrid-capture sequencing data for family size 1 data, showing a Pearson r of 0.91.

WGS data underwent pre-INVAR data processing, as described in Methods 4.7.2 and 4.7.3, except the minimum depth at a locus was set to 1, and patient-specific outlier-suppression was not used because loci with signal vs. loci without signal will only show allele fractions of 0 or 1, given an average depth of 0.6x.

4.6.9 Experimental spike-in dilution series

Plasma DNA from one patient with a total of 5,073 patient-specific variants was serially diluted 10-fold in a pool of plasma cfDNA from 11 healthy individuals (Seralab) to give a dilution series spanning 1-100,000x. Library preparation was performed, as described in Method 4.6.5, with 50ng input per dilution. The healthy control cfDNA pools were included as control samples for the determination of locus error rate to identify and exclude potential SNP loci.

To assess the sensitivity of INVAR with smaller panel sizes, the spike-in dilution series was downsampled to between 1 and 5,000 patient-specific loci. Given the relationship between tumour allele fraction and plasma mutation representation (Fig. 4.8), any smaller panel for INVAR should be based on clonal mutations with highest priority, with lower allele fractions included only if plasma sequencing data is sufficiently broad. Thus, we iteratively sampled the data with replacement from each of the dilution series sequencing libraries (with 50 iterations), then selected the top N mutations. The locus with the highest mutant allele fraction was the *BRAF* V600E mutation. After downsampling the number of loci, outlier-suppression (Method 4.7.5) was repeated on all samples except for the single *BRAF* V600E locus data.

4.6.10 Imaging

CT imaging was acquired as part of the standard of care for each patient of the stage IV melanoma cohort and were examined retrospectively. Slice thickness was 5 mm in all cases. All lesions with a largest diameter greater than 5 mm were outlined slice by slice on the CT images by an experienced operator, under the guidance of a radiologist, using custom software written in MATLAB (Mathworks, Natick, MA). The outlines were subsequently imported into LIFEx software in NifTI format for processing. Tumour volume was then reported by LIFEx as an output parameter from its texture based processing module.

4.7 Bioinformatics and Statistical Methods

4.7.1 Generation of patient-specific mutation lists

We generated patient-specific mutation lists for 64 patients with stage II-IV melanoma and stage I-IIIA lung cancer. A median of 625 (IQR 411-1076) and 388 (IQR 230-600) patient-specific mutations were identified per patient with melanoma and lung cancer, respectively (Figs. 4.26, 4.27, Table 4.3). These mutation lists were used both to design custom capture sequencing panels, and as input for the INVAR algorithm (Method 4.7.6).

For fresh frozen tumour biopsies, mutation calling was performed as described by Varela et al. [274]. For FFPE tumour biopsies, mutation calling was performed with Mutect2 with the default settings: –cosmic v77/cosmic.vcf and –dbsnp v147/dbsnp.vcf. To maximise the number of mutations retained, variants achieving Mutect2 pass (LUCID and AVAST-M samples) OR tumour LOD > 5.3 were retained (AVAST-M samples), with the additional filters:

- Buffy coat mutant allele fraction equals zero
- Mutation not in homologous region
- Mutation not at a multiallelic locus
- 1000 Genomes ALL and EUR frequency equals zero
- A minimum tumour depth of 5.
- For FFPE data in the melanoma cohort, the filter for C/A errors proposed by Costello et al. [284] was applied to suppress C/A artefacts arising from library preparation of FFPE samples.

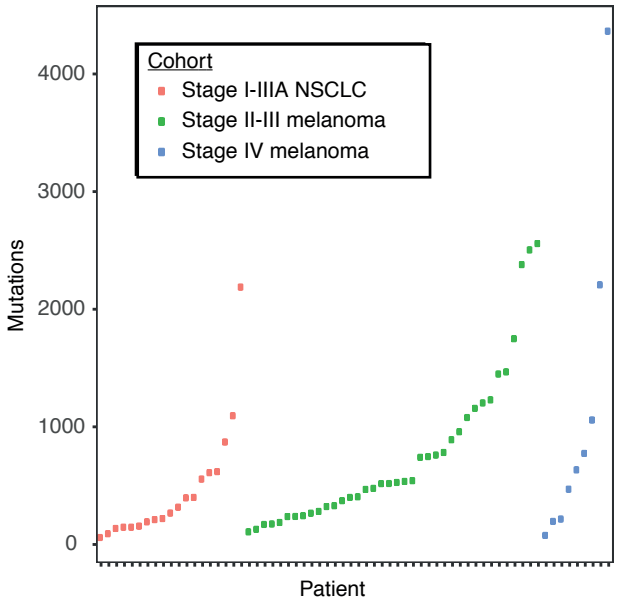


Fig. 4.26 **Individualised sequencing panel mutation profiles**
The number of somatic mutations per patient is shown, coloured by cancer type. Patients are ordered by increasing number of mutations per mutation list.

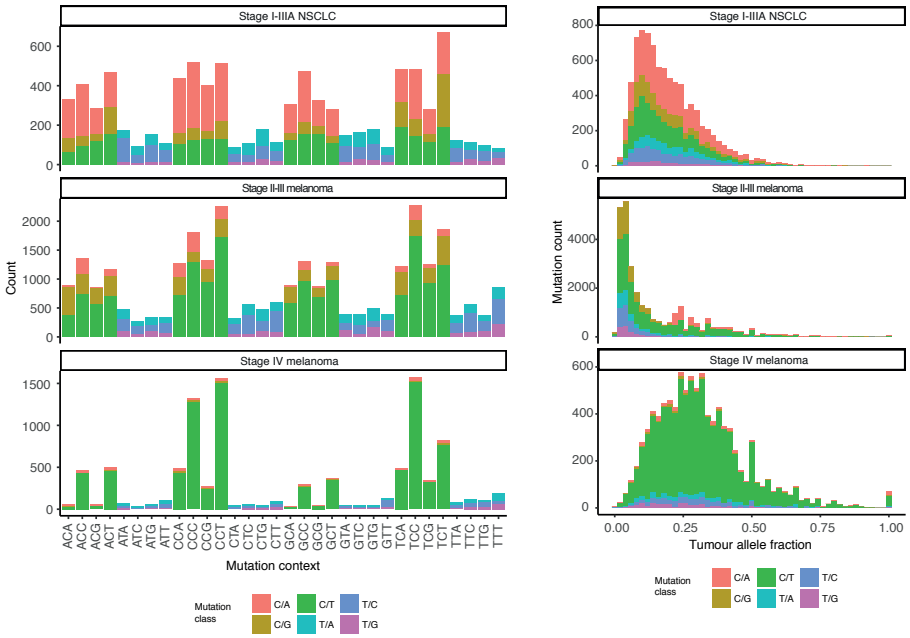


Fig. 4.27 **Tumour mutation profiles by trinucleotide context**
Mutation counts by trinucleotide context, coloured by mutation class. The allele fraction distribution is shown for each cohort. The Stage I-IIIA NSCLC and stage II-III melanoma cohorts were called from FFPE material, and the stage IV melanoma cohort was called from fresh frozen tumour material.

4.7.2 Data processing and error-suppression

Cutadapt v1.9.1 was used to remove known 5' and 3' adaptor sequences specified in a separate FASTA of adaptor sequences. Trimmed FASTQ files were aligned to the UCSC hg19 genome using BWA-mem v0.7.13 with a seed length of 19. Error-suppression was carried out on ThruPLEX Tag-seq library BAM files using CONNOR [285]. The consensus frequency threshold -f was set as 0.9 (90% consensus), and the minimum family size threshold -s was varied between 2 and 5 for characterisation of error rates (Fig. 4.28). For custom capture and exome sequencing data, a minimum family size of 2 was used. For sWGS, a minimum family size of 1 was used.

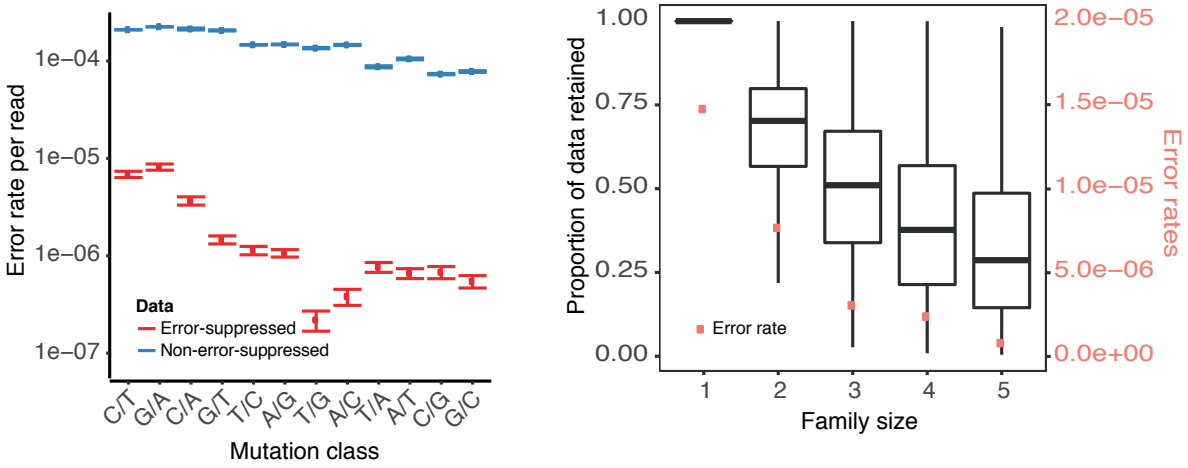


Fig. 4.28 **Background error rates with and without error-suppression**
Error suppression was performed using CONNOR [285] with different minimum family size requirement settings, ranging from 1 to 5. To assess background error rate, 10 bp either side of patient-specific loci were used, excluding the patient-specific locus itself ('near-target'). Background error rates were calculated by aggregating all non-reference bases across all bases considered. Left panel, error suppressed and non-error suppressed background error rates. Right panel, overall background error rates resulting from different family size requirements, and proportion of read families retained with each setting.

To leverage signal across multiple time points, error-suppressed BAM files can be combined using 'samtools view -ubS - | samtools sort -' prior to further data processing. In the early-stage melanoma cohort (AVAST-M), where samples were available at

both 3 and 6 month time points post-surgery, BAM files were run through the INVAR pipeline both individually (Table 4.4) and after merging (Figs. 4.19, 4.20).

4.7.3 Data filtering

The INVAR pipeline takes error-suppressed BAM files, a BED file of patient-specific loci, and a CSV file indicating the tumour allele fraction of each mutation and which patient it belongs to. It is optimised for a cluster running Slurm. The workflow is shown in Fig. 4.29. Briefly, the pipeline assesses wild-type and mutant reads at patient-specific loci in all samples, and this data is annotated with trinucleotide error rate, locus error rate, which patient the mutation belongs to, tumour allele fraction, fragment size, presence in both F and R reads, and whether the signal at that locus is an outlier relative to all other patient-specific loci in that sample (Method 4.7.5). Following data annotation, signal is aggregated across all patient-specific loci in that sample to generate both an INVAR score, equivalent to a specificity (Method 4.7.6) and an integrated mutant allele fraction (IMAF), which is the estimate of ctDNA level based on the INVAR algorithm (Method 4.7.8).

SAMtools mpileup 1.3.1 was used at patient-specific loci based on a BED file of muations, with the following settings: –ff UNMAP, –q 40 (mapping quality), –Q 20 (base quality), –x, –d 10,000, then multiallelic calls were split using BCFtools 1.3.1. Next, all TSV files were annotated with 1,000 Genomes SNP data, COSMIC data, and trinucleotide context using a custom Python script. Output files are then concatenated, compressed, and read into R. First, based on prior knowledge from tumour sequencing data, all loci were annotated per patient with being either: patient-specific (present in patient's tumour) or non-patient-specific (not present in patient's tumour, or individual does not have cancer). Since each non-patient-specific sample contains the loci from multiple patients, every non-patient-specific sample may control for all other patients

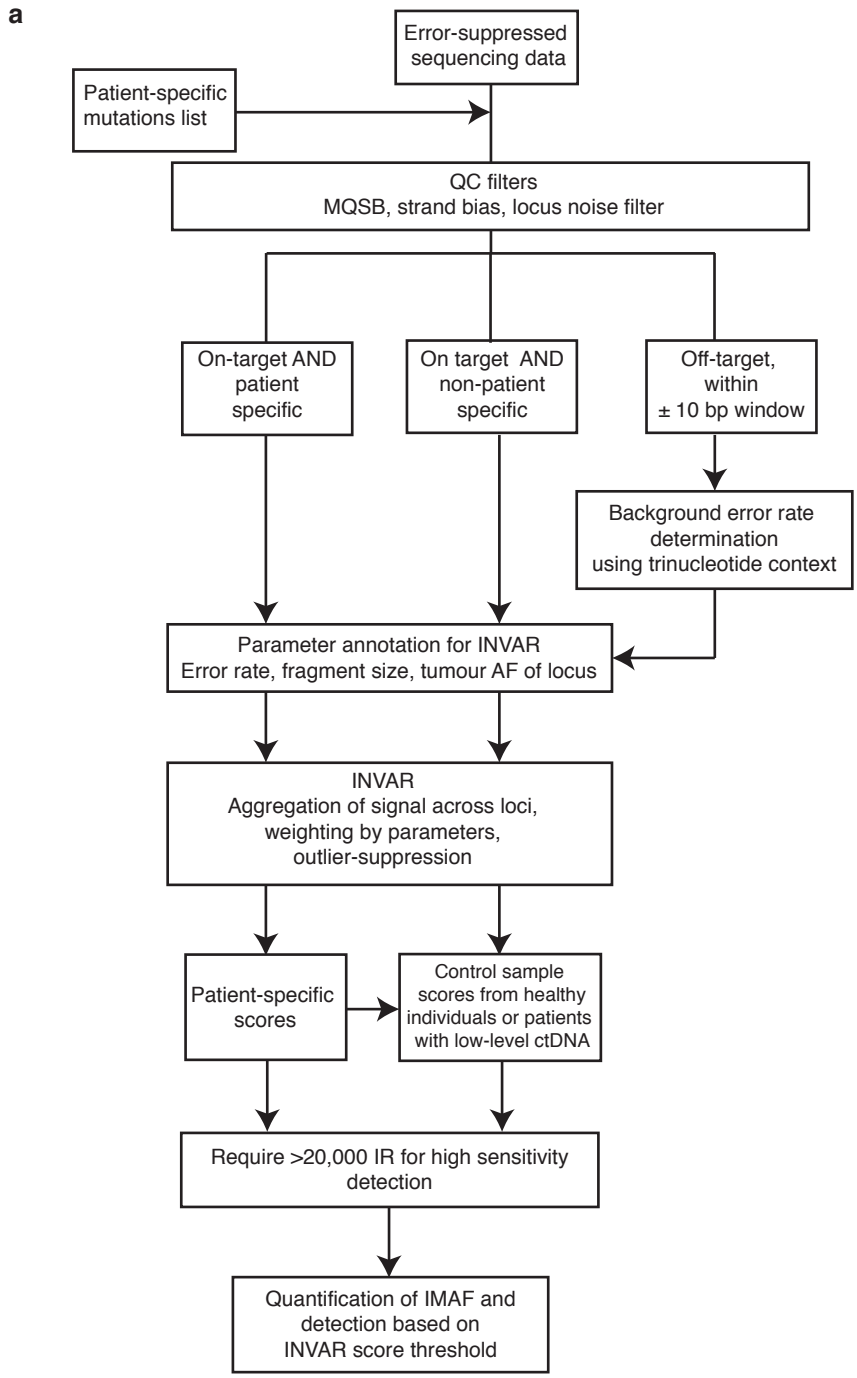


Fig. 4.29 **Overview of the INVAR workflow**
INVAR utilises plasma sequencing data and requires a list of patient-specific mutations, which may be derived from tumour or plasma sequencing. Filters are applied to sequencing data, then the data is split into: patient-specific (locus belonging to that patient), non-patient-specific (locus not belonging to that patient), and near-target (bases within 10 bp of all patient-specific loci). Patient-specific and non-patient-specific data are then annotated with features that influence the probability of observing a real mutation. Outlier-suppression is applied to identify mutant signal inconsistent with the overall level of patient-specific signal. Next, signal is aggregated across all loci, taking into account annotated features, to generate an INVAR score per sample. Based on non-patient-specific samples, an INVAR score threshold is determined based on ROC analysis for each cohort.

analysed with the same sequencing panel or method (excluding loci that are shared between individuals). Next, the following filters were applied to both patient-specific and non-patient-specific data:

- Data points were excluded if MQSB < 0.01 (mapping quality / strand bias).
- Multi-allelic sites were identified, and were blacklisted if 3 different alternate alleles were observed with error-suppressed read families in the dataset. Loci with 2 separate alternate alleles observed in the dataset were only excluded if there were more than 2 error-suppressed reads of the minor alternate allele.
- Loci were blacklisted on the basis of strand bias if they showed a ratio between F and R mutant reads <0.1 or >10, with mutant signal in at least three separate samples.
- Loci that showed mutant signal in >10% of the control samples were excluded (Fig. 4.30). When determining the error rate per locus, all other patients who were genotyped as negative for that mutation may serve as controls for the patient of interest, since 99.8% of mutations are private to each patient.
- Mutation signal had to be represented in both the F and R read of that read pair (Fig. 4.31). This both serves to reduce sequencing error and causes a size-selection for fragments, retaining fragments <300bp as PE150 sequencing was performed (only mutant signal in the overlapping region of the F and R read can be retained). The resulting error-suppression is analogous to tools that merge paired-end reads [286].

The effects of combining each of these filters on the background error rate are shown in Fig. 4.31.

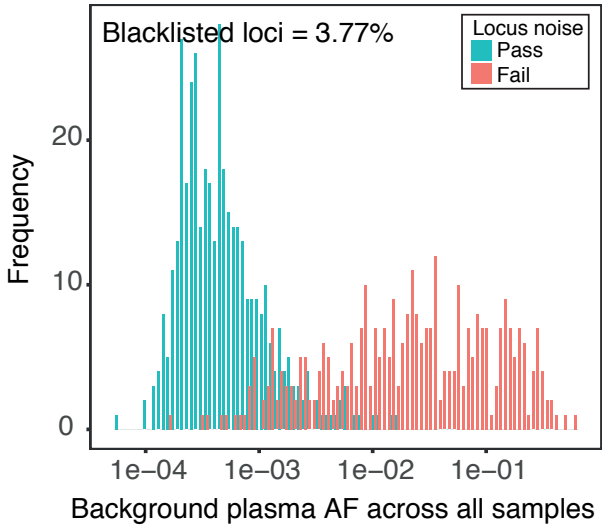


Fig. 4.30 **Allele fraction distribution of loci removed by the locus-noise filter**
For each patient-specific locus, the error rate was determined in all other patient samples. Loci were blacklisted if they showed signal in >10% of control samples, or had an overall allele fraction of >1% in control samples. The distribution of allele fractions of blacklisted and non-blacklisted loci that had non-zero allele fraction are shown.

4.7.4 Feature annotation

After data filtering, data was annotated with both locus-specific error rate and trinucleotide error rate. Locus-specific error rate is calculated as the background error rate at that locus in control samples for the patients-specific mutation. Since the estimation of locus-specific error rate is limited by the number of control samples and cfDNA molecules at that locus, we also assessed trinucleotide error rate. Trinucleotide error rates were determined from the region up to 10bp either side of every patient-specific loci (excluding patient-specific loci), and data was pooled across all similar trinucleotide contexts. After pooling data in this manner, a median of 3.0 x 10^8 informative reads (or deduplicated reads) per trinucleotide context were analysed. Trinucleotide error rate was calculated as a mismatch rate for each specific mutation context. If a trinucleotide context had zero mutant deduplicated reads, the error rate was set to the reciprocal of the number of IR/deduplicated reads in that context.

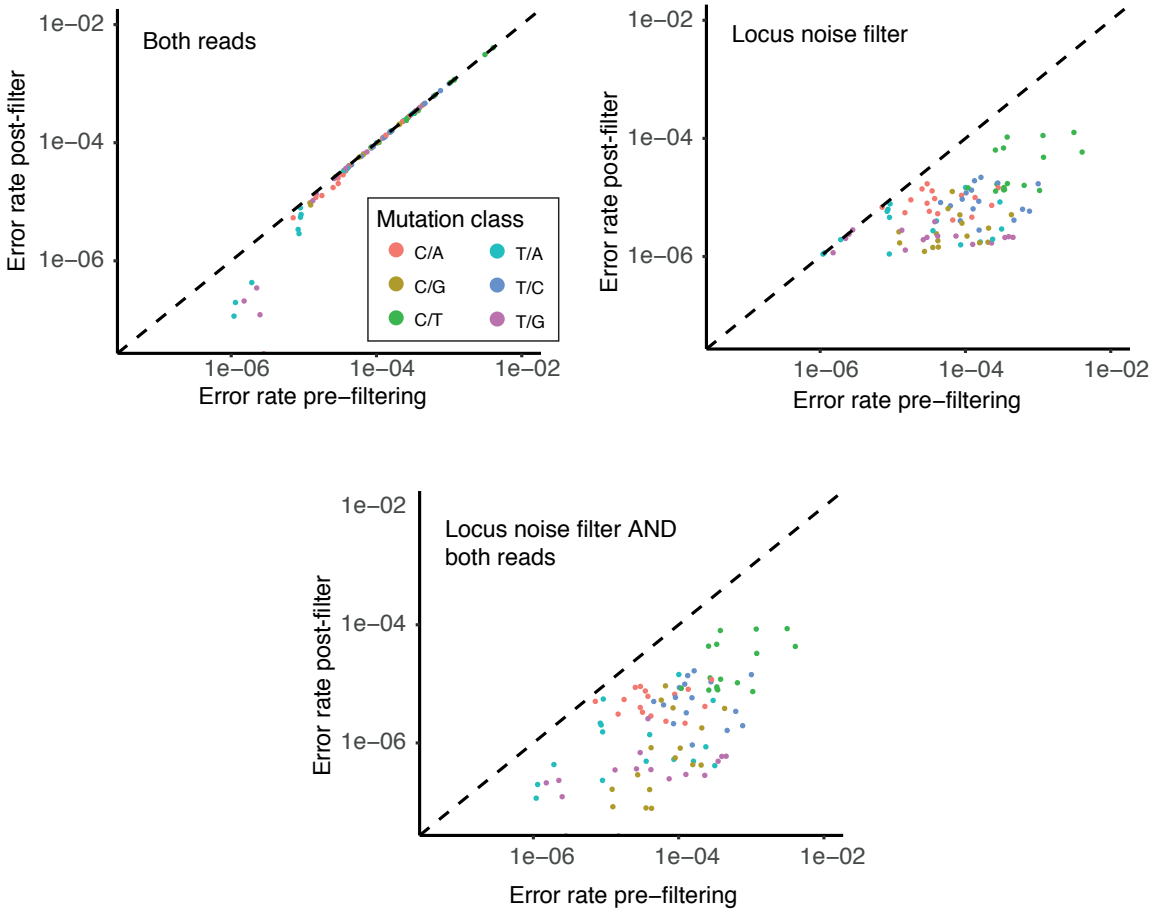


Fig. 4.31 **Effect of each noise filter on background-error rate**
For each trinucleotide, background error rates are plotted before and after each background error filter, i.e. requiring signal in both reads, and a locus noise filter (Method 4.7.3). Each filter produced a separate effect: requiring signal to be present in both forward and reverse reads only reduced error rates of trinucleotides with error < 5 x 10^{-5}.

In addition, each data point was annotated with the cfDNA fragment size of that read using a custom Python script. Then, to eliminate outlier signal that is not consistent with the remainder of that patient's loci, we performed patient-specific outlier suppression (Method 4.7.5, below).

4.7.5 Patient-specific outlier-suppression

Patient-specific sequencing data consists of informative reads at multiple known patient-specific loci, providing the opportunity to compare mutant allele fractions across loci as a means of error-suppression. The distribution of signal across loci potentially allows for the identification of noisy loci not consistent with the overall signal distribution. Each locus was tested for the probability of having observed mutant reads given the average signal across all loci (Figs. 4.32, 4.33). Loci observed with significantly greater signal than the remainder of the loci might be due to noise at that locus, contamination, or a mis-genotyped SNP locus. The possibility of a mis-genotyped SNP becomes increasingly likely when a large number of mutation loci are targeted by INVAR.

For each sample, the IMAF was determined across all loci passing pre-INVAR data processing filters with mutant allele fraction at that locus of <0.25; loci with signal >0.25 mutant allele fraction were not included in the calculation because (i) in the residual disease setting, loci would not be expected to have such high mutant allele fractions (unless they are mis-genotyped SNPs), and (ii) if the true IMAF of a sample is >0.25, when a large number of loci are tested, they will show a distribution of allele fractions such that detection is supported by having many low allele fraction loci with signal. Based on the ctDNA level of the sample, the binomial probability of observing each individual locus given the IMAF of that sample was calculated. Loci with a Bonferroni corrected P-value <0.05 (corrected for the number of loci interrogated) were excluded in that sample, thereby suppressing outliers. As a result of outlier-suppression, background noise was reduced 2.3-fold in control samples, while retaining 94.7% of signal in patient samples (Fig. 4.33).

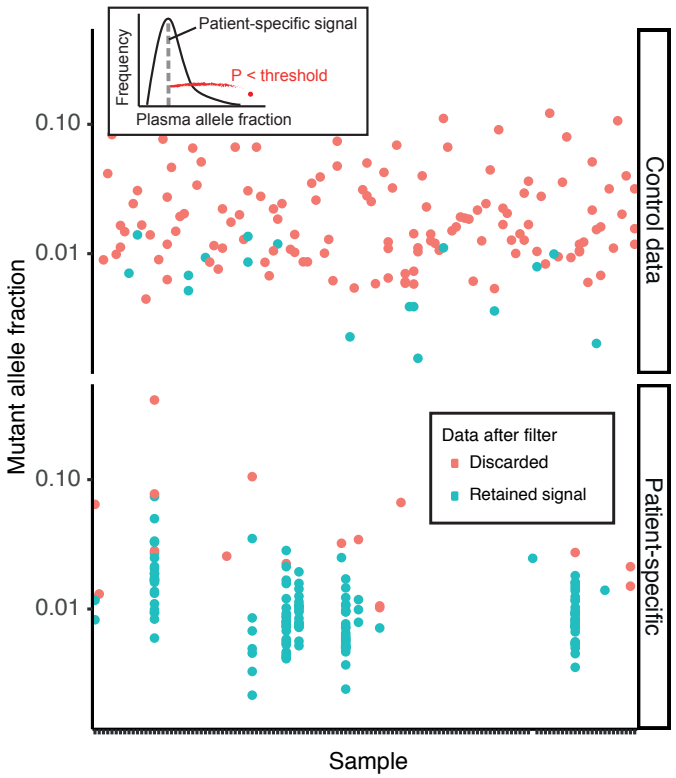


Fig. 4.32 **Effect of outlier-suppression on mutant signal**
Effect of outlier-suppression in early-stage NSCLC cohort. Loci observed with significantly greater signal than the remainder of the loci of that patient might be due to noise at that locus, contamination, or a mis-genotyped SNP locus and were removed (red, see Method 4.7.5).

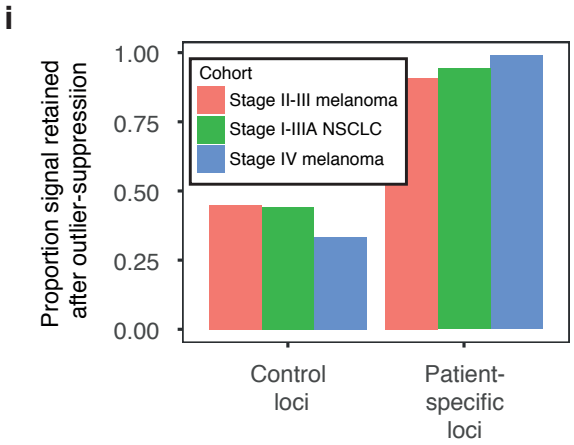


Fig. 4.33 **Summary of effect of outlier-suppression on all three cohorts**
Following patient-specific outlier-suppression, background noise was reduced 2.3-fold in control samples, while retaining 94.7% of signal in patient samples.

4.7.6 Statistical model for detection

We developed a statistical method to model the number of mutant reads at multiple patient-specific loci, incorporating prior information available from patient-specific sequencing, such as the background error of that trinucleotide context, the tumour allele fraction at that locus, and fragment length. This approach aggregates signal across multiple patient-specific mutations after error-suppression (Method 4.7.2). In this model, we test the significance of the number of mutant reads at each locus given the trinucleotide error rate of that context. Trinucleotide error rates were used instead of locus-specific error rates in order to determine an accurate estimation of background error rates to 10^{-7} (Fig. 4.6). Locus-specific error rates were previously used to identify loci with recurrent signal in control samples (Method 4.7.3).

Tumour allele fractions and trinucleotide error rates were considered as follows: Denote AF_i as the tumour mutant allele fraction at locus i, e_i as the background error in the context of locus i, and let p be the integrated mutant allele fraction of a sample (IMAF, estimated using an expectation maximisation algorithm in Method 4.7.8). A random read at locus i can be observed to be mutant either if it arose from a mutant molecule, or an incorrectly sequenced wild type DNA molecule; this event occurs with probability q_i:

$$q_i = AF_i \cdot (1 - e_i) \cdot p + (1 - AF_i) \cdot e_i \cdot p + e_i \cdot (1 - p)$$

Testing for the presence of ctDNA is now equivalent to testing the statistical hypothesis $H_0 : p = 0$. Assuming the number of observed mutant reads is independent between loci, the following likelihood function can be produced:

$$L(p;M,AF,e) = \prod_{i=1}^{n}\prod_{j=1}^{R_i} {q_i}^{M_{ij}}(1-q_i)^{1-M_{ij}},$$

where M_{ij} is the indicator for mutation in read j of locus i, and R_i is the number of reads in locus i. The above method allows weighting of signal by tumour allele fraction, which we confirm influences plasma mutation representation in patient samples with early stage and advanced disease (Fig. 4.8), and in the spike-in dilution series from one patient (Fig. 4.9, Method 4.6.9).

Each sequencing read provides fragment size information, which may be used to separate mutant from wild-type molecules and produce an enrichment in ctDNA (Fig. 4.7). Probability weighing was preferred over size selection to avoid allelic loss at ultra-low allele fractions, suggested by Fan et al. [208] in the non-invasive prenatal testing setting. Therefore, read length information can also be incorporated into the likelihood. The read length distribution of mutant and wild-type fragments was estimated in Methods 4.7.7. This approach is in contrast to size-selection and may be considered as a size-weighting step alongside tumour AF weighting that was performed above. Fragment sizes for each sequencing read may be incorporated to the INVAR algorithm. To do so, let L_{ij} be the length of read i of locus j. The likelihood can be written as:

$$L(p;M,L,AF,e) = \prod_{i=1}^{n}\prod_{j=1}^{R_i} P(M_{ij},L_{ij}|e,AF,p).$$

Assuming that given the source of the read (mutant or wild-type DNA), the read length, and mutation status are independent, we can factor the likelihood as follows:

$$L(p;M,L,AF,e) = \prod_{i=1}^{n}\prod_{j=1}^{R_i} P(m_{ij},l_{ij}|z_{ij}=0)\cdot P(z_{ij}=0) + P(m_{ij},l_{ij}|z_{ij}=1)\cdot P(z_{ij}=1) =$$
$$\prod_{i=1}^{n}\prod_{j=1}^{R_i} P(m_{ij}|z_{ij}=0)\cdot p^0(l_{ij})\cdot(1-p) + P(m_{ij}|z_{ij}=1)\cdot p^1(l_{ij})\cdot p =$$
$$\prod_{i=1}^{n}\prod_{j=1}^{R_i} {e_i}^{m_{ij}}\cdot(1-e_i)^{1-m_{ij}}\cdot p^0(l_{ij})\cdot(1-p) + g_i^{m_{ij}}\cdot(1-g_i)^{1-m_{ij}}\cdot p^1(l_{ij})\cdot p,$$

where z_{ij} is the indicator that read j of locus i came from ctDNA, $p^k(l_{ij}) = P(l_{ij}|z_{ij}=k)$, and $g_i = AF_i\cdot(1-e_i) + (1-AF_i)\cdot e_i$. The above method weights the signal based on both fragment length of mutant and wild-type reads, though in this implementation of INVAR, we set the weight of all wild-type size bins to be equal, thereby neglecting size information from wild-type reads. Lastly, a score is generated for each sample through aggregation of signal across all patient-specific loci in that sample using the Generalized Likelihood Ratio test (GLRT). The GLRT directly compares the likelihood under the null hypothesis against the likelihood under the maximum likelihood estimate of p (ctDNA fraction in plasma):

$$\lambda(p_0) = \frac{L(p_0;M,L,AF,e)}{L(\hat{p};M,L,AF,e)}.$$

The higher the value of the likelihood ratio, the greater evidence for ctDNA presence in a sample. Classification of samples was performed based on comparison of likelihood ratios between patient and control samples (Method 4.7.9).

4.7.7 Estimation of the read length distribution

Size-weighting with INVAR depends on first having a known distribution of sizes of mutant and wild-type reads against which to perform weighting. In order to estimate

2402 the read length distribution with the greatest accuracy, we used all wild type and mutant
2403 reads from healthy samples and patients from each of the cohorts, and used kernel
2404 density estimation to smooth the respective probabilities.

2405 The size distributions and enrichment ratios per bin for each of the studied cohorts are
2406 shown in Fig. 4.7. We demonstrated that the early stage cohorts were not significantly
2407 different in size profile, whereas the advanced melanoma cohort had a significantly
2408 greater proportion of di-nucleosomal fragments despite downsampling of data to a
2409 similar number of reads (Fig. 4.34). Thus, data from both early stage cohorts were
2410 pooled to generate a prior distribution of the size of mutant and wild-type fragments,
2411 and data was smoothed with a Gaussian kernel with a default setting of 0.25 (Fig. 4.35).

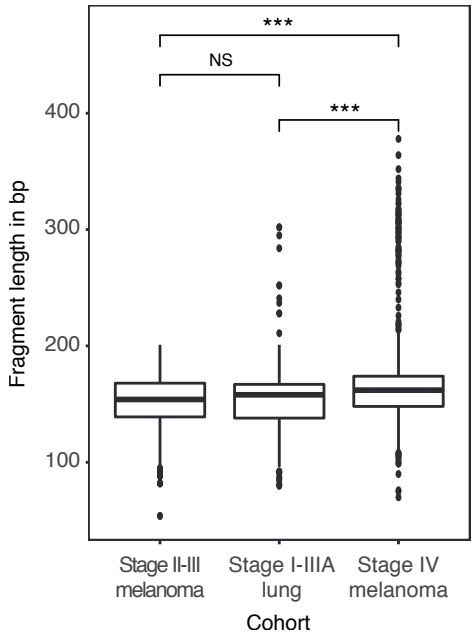


Fig. 4.34 **Comparison of mutant fragment distribution between cohorts**
Fragment size distributions were compared using a two-tailed Wilcoxon rank test. Data were first downsampled to the same number of reads. Stage IV melanoma patients show a significantly longer mutant size profile compared to the early stage cohorts. NS, not significant; *** = P < 0.0001.

2412 To estimate the probability that a read is of length l, given that the cell of origin of
2413 wild type, $P(L = l|z = 0)$, we used all of the wild-type reads from each pooled data set.

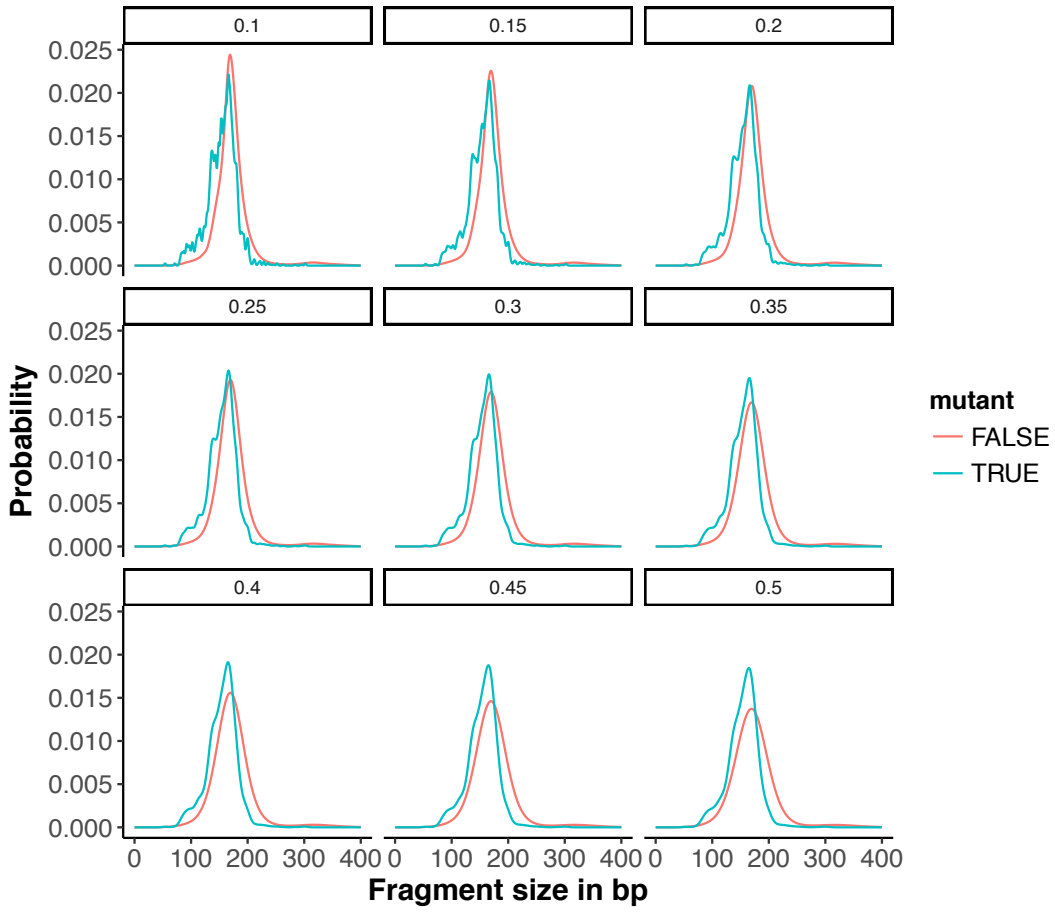


Fig. 4.35 **Effect of smoothing size profile data**
In order to weight fragment sizes, mutant and wild-type size profiles were characterised and smoothed to varying extents (Methods 4.7.7). For each level of smoothing, the distribution of mutant and wild-type fragment sizes is shown.

For both of the data sets, we used the R function "density", with a Gaussian kernel, to smooth the estimated probabilities, and obtained a density estimate $\hat{f}(l|Z=z)$. Finally, to estimate $P(L=l|Z=z)$, we integrated the respective density: $P(L=l|Z=z) = \int_{l-0.5}^{l+0.5} \hat{f}(t|Z=z)dt$.

Smoothing the size distribution estimates is important in data sets where data is sparse to avoid assigning too large a weight to any given mutant fragment.

4.7.8 Calculation of IMAF

In this section we derive an Expectation Maximization (EM) algorithm to estimate IMAF (integrated mutant allele fraction). If we treat the tumor of origin z_{ij}, as a latent variable, and assume that is known, the joint likelihood of Z, M ($M_i j$ is the indicator for a mutation in read j of locus i), L ($L_i j$ is the length of read i of locus j), AF (AF_i is the mutant AF at locus i), e (e_i is the background error in the context of locus i) can be written as:

$$L(p;Z,M,L,AF,e) = \prod_{i=1}^{n}\prod_{j=1}^{R_i}\left[{e_i}^{m_{ij}}\cdot(1-e_i)^{1-m_{ij}}\cdot p^0(l_{ij})\cdot(1-p)\right]^{1-z_{ij}}\cdot \left[g_i^{m_{ij}}\cdot(1-g_i)^{1-m_{ij}}\cdot p^1(l_{ij})\cdot p\right]^{z_{ij}} \tag{4.1}$$

We can now use EM to find a maximum likelihood estimate for p, and taking the expectation of the likelihood with respect to z_{ij}. The log-likelihood is linear in z_{ij}, so taking the expectation of the likelihood amounts simply to replacing the Z_{ij} with their expectation at stage l, $z_{ij}^l = E(z_{ij}|m_{ij},l_{ij},p_l)$, where p_l is the best estimate of p at iteration l. An estimate for p_l is obtained by taking the derivative with respect to p_l and equating it to zero:

$$p_l = \frac{\sum_{i=1}^{n}\sum_{j=1}^{R_i} z_{ij}^{l}}{\sum_{i=1}^{n} R_i},$$

The above is simply the expected proportion of reads from ctDNA at stage l.

Bayes' theorem can be used to compute z_{ij}^{l}:

$$z_{ij}^{l} = P(z_{ij}=1|m_{ij},l_{ij},p_l) = \frac{P(m_{ij}|z_{ij}=1)\cdot p^1(l_{ij})\cdot p}{P(m_{ij}|z_{ij}=1)\cdot p^1(l_{ij})\cdot p + P(m_{ij}|z_{ij}=0)\cdot p^0(l_{ij})\cdot(1-p)}.$$

By substituting the respective probabilities we obtain:

$$z_{ij}^{l} = \frac{g_i^{m_{ij}}(1-g_i)^{1-m_{ij}}p^1(l_{ij})p}{g_i^{m_{ij}}(1-g_i)^{1-m_{ij}}p^1(l_{ij})p + e_i^{m_{ij}}(1-e_i)^{1-m_{ij}}p^0(l_{ij})(1-p)}$$

The algorithm proceeds by alternating the maximization of p, and taking the expectation of the z_{ij}s.

4.7.9 Detection classification based on INVAR scores

For each sample, the likelihood ratio (LR) was determined, as above. To accurately assess the background LR distribution, reads from each control sample were iteratively resampled with replacement and the GLRT script run. To convert each LR into an INVAR score for each sample, the proportion of controls with an equal or lower LR was calculated, making the INVAR score analogous to a specificity value.

Using control samples, the cut-off for INVAR score was determined using ROC analysis using the 'OptimalCutpoints' package in R [287] . To minimise the risk of any patient-specific contamination of signal at non-patient-specific control loci (through *de novo* mutations overlapping with patient-specific sites), only patient samples with

patient-specific IMAF <1% were used as controls for determination of the cut-point. ROC analysis gave INVAR score thresholds varying between 0.92-1.00 for the above cohorts, which are provided in Table 4.5. Samples with INVAR scores greater than their respective threshold were classified as detected.

4.7.10 Calculation of informative reads (IR)

The number of informative reads (IR) for a sample is the product of the number of mutation targeted (i.e. length of the mutation list) and the number of haploid genomes analysed by sequencing (hGA, equivalent to the deduplicated coverage following read-collapsing). Thus, the limit of detection for every sample can be calculated based on 1/IR (with adjustment for sampling mutant molecules based on binomial probabilities).

For non-detected samples, the 1/IR value provides an estimate for the upper limit of ctDNA in that sample; this allows quantification of samples even if no mutant molecules are present, and is utilised in Fig. 4.23 to demonstrate quantification to 10^{-4} using sWGS data. Also, samples with limited sensitivity can be identified and classified as a 'low-sensitivity' or 'non-evaluable' group, where the INVAR algorithm is limited by the number of IR. In this study, we aimed to quantify ctDNA with sensitivity greater than other methods, and so we classified samples with non-detected ctDNA with <20,000 IR as low-sensitivity and thus non-evaluable. Across the cohorts in this study, 7 patients were non-evaluable with these criteria.

4.7.11 Survival analysis for resected stage II-III melanoma cohort

Disease-free interval (DFI), and overall survival were calculated from the date of randomisation of the AVAST-M trial to the date of first recurrence or date of death, respectively [266]. Kaplan-Meier analysis was used to generate survival curves for differences between DFI and OS in patients with detected ctDNA vs. non-detected levels and compared using a Cox proportional hazards model to obtain hazard ratios and 95% CIs.

4.7.12 Estimated detection rates with fewer informative reads

Based on the IMAFs of detected samples, detection rates can be estimated where fewer IR were achieved with a perfectly sensitive assay. For a given number of IR, the value of p that would be detected with 95% sensitivity can be determined as follows:

$$p = 1 - e^{\log(1\text{-}0.95)/\mathrm{IR}}$$

Thus, for each entry in a vector of IR values (10^2, 10^3 … 10^7), the detection rates for cancer were calculated per cohort, and are plotted in Fig. 4.21. The maximum value of the vector of IR values was set to be larger than the maximum number of IR per sample in that cohort, rounded to the nearest order of magnitude. For the stage II-III melanoma patients, detection was defined as the sensitivity for patients who relapsed within 5 years. Linear regression was used to calculate R^2 values for each cohort.

4.8 Acknowledgements

The authors would like to thank Catherine Thorbinson, Alex Azevedo, Vicky Senior, Jenniffer Castedo, Amanda Stone, Ellen Mosseley, Tim Young, Julia Knight, Meryl Griffiths, Beverley Haynes, Amy Gladwell, Neera Maroo, Amanda Walker and Michele Bianchi from the MelResist, LUCID and AVAST-M study groups, and the Cambridge Cancer Trial Centre, Addenbrookes Hospital and The Royal Papworth Hospital. We would like to acknowledge the support of The University of Cambridge, and Cancer Research UK (grant numbers A11906, A20240, and C2195/A8466). The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n.337905. RCR and DMR are supported, in part, by the NIHR BRC Cambridge.

4.9 Data and Materials Availability

Raw sequencing data will be made available at the European Genome-phenome archive, accession number EGAS00001002959. The INVAR pipeline will be made publicly accessible at http://www.bitbucket.org/nrlab/invar.

4.10 Supplementary Figures

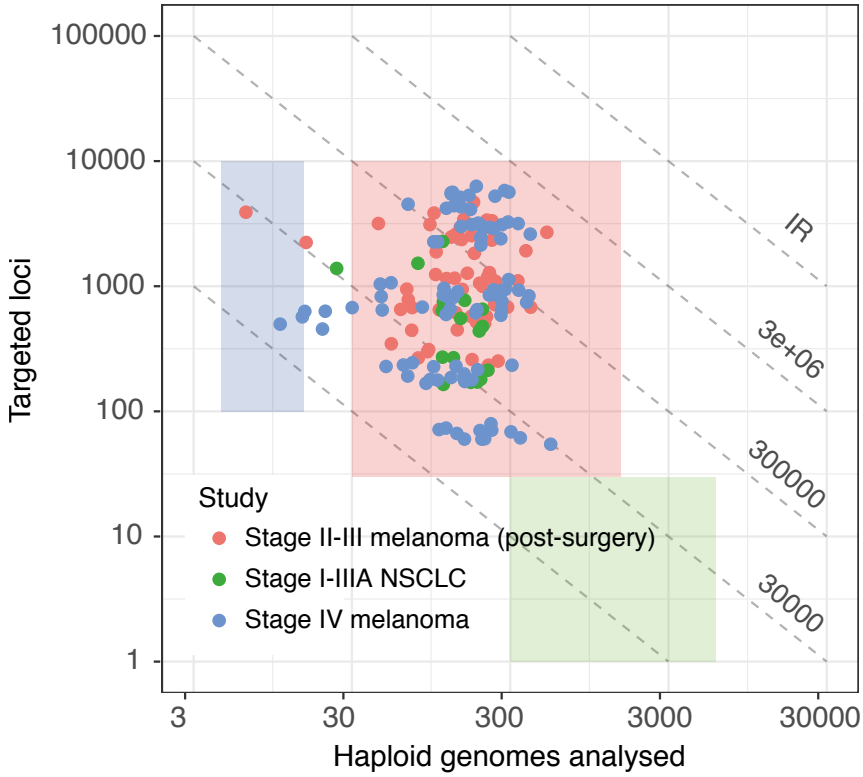


Fig. 4.36 **Number of hGA vs. mutations targeted in the conventional input setting** The shaded red box indicates the working point of INVAR using millilitres of plasma. Samples are coloured by cohort. Dashed diagonal lines indicate the number of IR achieved for a given number of hGA and targeted loci. hGA, haploid genomes analysed; IR, informative reads.

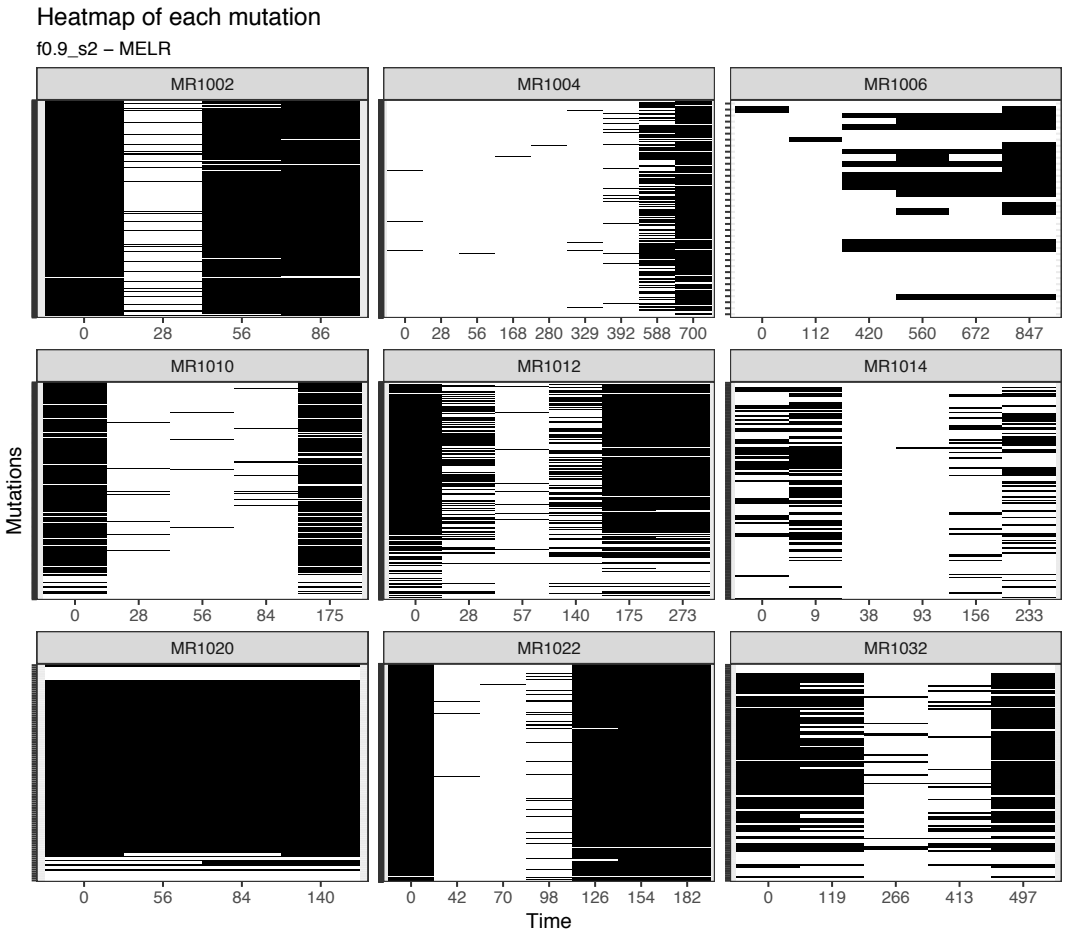


Fig. 4.37 **Heatmap of individual mutations over time**
For each patient with stage IV melanoma, a heatmap showing their mutations over time is plotted. Each line represents one patient-specific mutation. Mutations are either filled or not filled, based on the mutation being observed in plasma or not observed, respectively. The effect of sampling error for individual mutations at low levels of ctDNA can be observed.

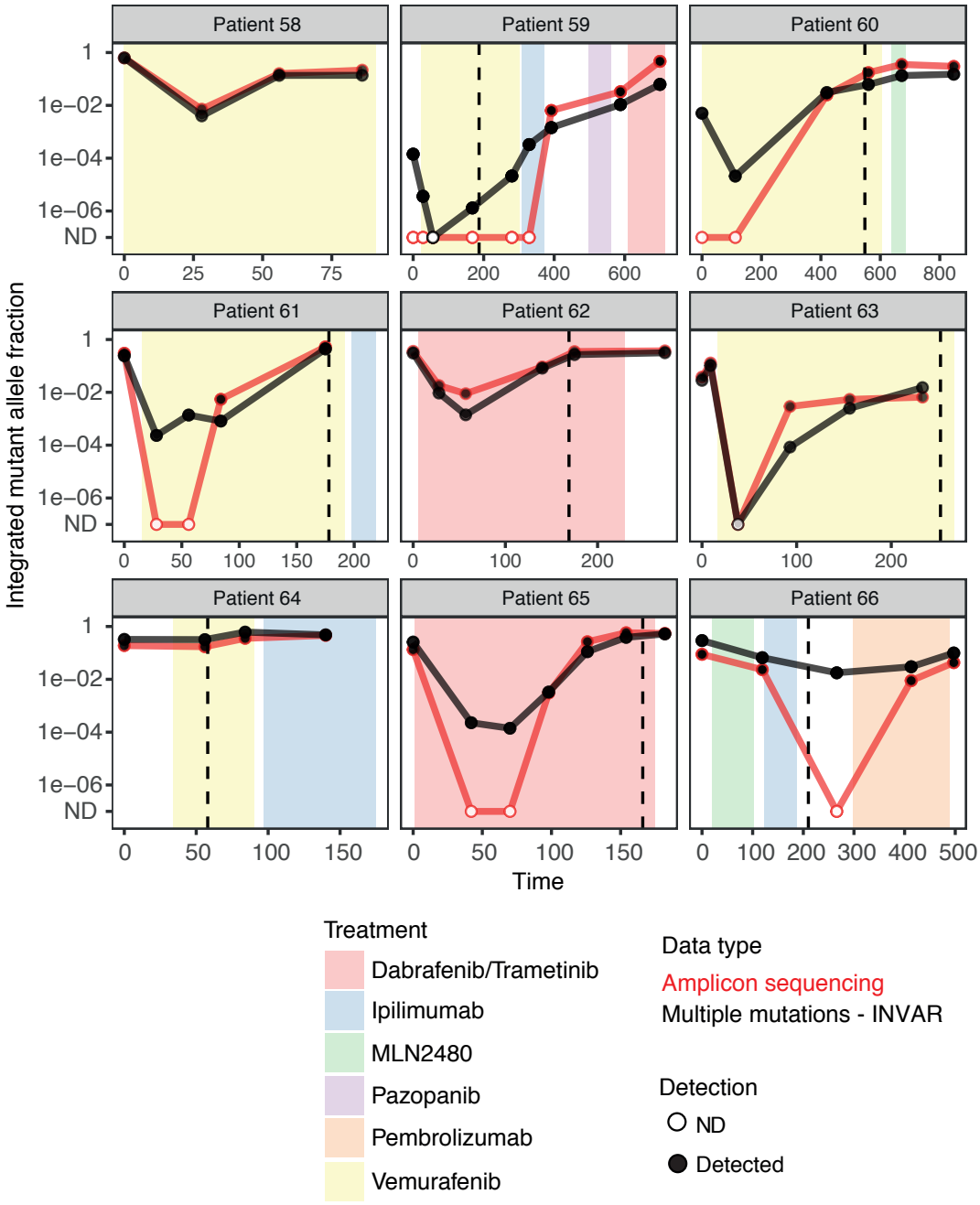


Fig. 4.38 **Comparison of INVAR IMAF against amplicon sequencing mutant allele fraction in advanced stage melanoma patients**
Amplicon sequencing was performed on a subset of samples with TAm-Seq [29], targeting one mutation per patient, either a *BRAF* or *NRAS* mutation. A Pearson correlation of 0.85 was observed between the two methods.

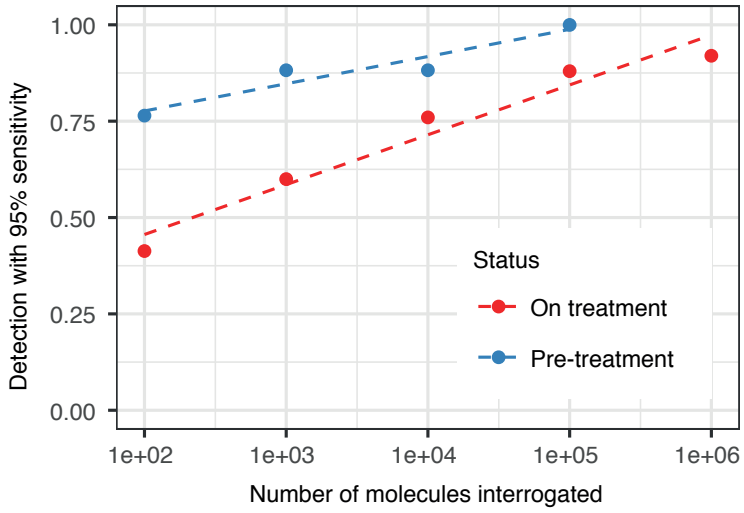


Fig. 4.39 **Detection rate with varying IR in melanoma pre- and post-treatment initiation** We estimated the detection rates of ctDNA for different levels of IR (Method 4.7.12). We observe a linear relationship ($R^2 = 0.95$) between the number of IR and detection rate in the baseline samples of the stage IV melanoma cohort (blue). ctDNA was detected in 100% of these samples with 10^5 IR. Detection of ctDNA in follow-up samples would require at least 10^6 IR (red).

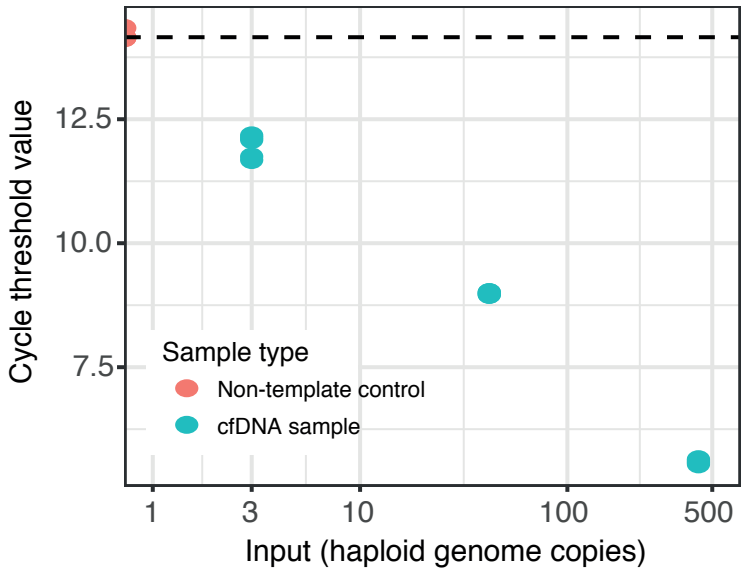


Fig. 4.40 **Library preparation from individual genome copies** Preparation of DNA libraries from restricted inputs. Libraries were prepared from diluted samples ranging from 3 – 500 haploid genome copies. Cycle threshold values were compared to non-template control (red).

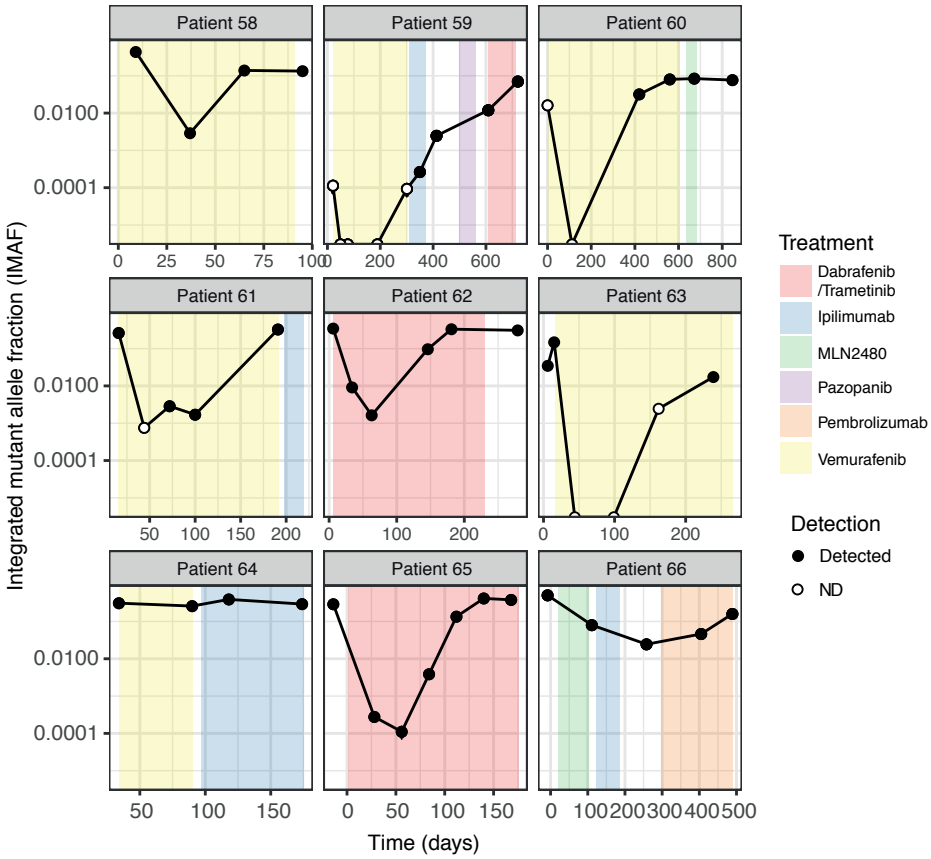


Fig. 4.41 **Monitoring stage IV melanoma patients using data downsampled *in silico* to 10 hGA**

ctDNA was monitored to 2 x 10^{-4} IMAF using 10 hGA, which is equivalent to 10 genome copies, which we estimate is the cfDNA yield obtainable from a droplet of blood. Filled circles indicate detection.

4.11 Tables

Chromosome	Position	Reference	Alternate	Gene	Depth	Tumour allele fraction	Patient	Study	Cancer type	Custom capture panel number	Mutation class
chr1	13418	G	A	DDX11L1	189	0.101	AM445	AVASTM	melanoma	7	G/A
chr1	664834	T	G	RP11-206L1C	59	0.102	AM445	AVASTM	melanoma	7	T/G
chr1	2100638	T	A	PRKCZ	49	0.02	AM445	AVASTM	melanoma	7	T/A
chr1	7887590	T	G	PER3	100	0.01	AM445	AVASTM	melanoma	7	T/G
chr1	11128701	C	T	EXOSC10	39	0.051	AM445	AVASTM	melanoma	7	C/T
chr1	12907777	T	C	HNRNPCL1	40	0.375	AM445	AVASTM	melanoma	7	T/C
chr1	12939548	G	C	PRAMEF4	47	0.043	AM445	AVASTM	melanoma	7	G/C
chr1	13036654	C	T	PRAMEF22	39	0.538	AM445	AVASTM	melanoma	7	C/T
chr1	16892540	A	G	NBPF1	532	0.026	AM445	AVASTM	melanoma	7	A/G
chr1	17084707	T	C	MST1L	41	0.293	AM445	AVASTM	melanoma	7	T/C
chr1	17085485	G	A	MST1L	189	0.085	AM445	AVASTM	melanoma	7	G/A
chr1	20659355	G	C	VWA5B1	42	0.024	AM445	AVASTM	melanoma	7	G/C
chr1	24449903	C	T	IL22RA1	48	0.021	AM445	AVASTM	melanoma	7	C/T
chr1	26127578	T	A	SEPN1	69	0.014	AM445	AVASTM	melanoma	7	T/A
chr1	28573129	C	T	RP5-1092A3.	105	0.009524	AM445	AVASTM	melanoma	7	C/T
chr1	29424394	G	C	EPB41	68	0.015	AM445	AVASTM	melanoma	7	G/C
chr1	32120347	G	C	COL16A1	60	0.017	AM445	AVASTM	melanoma	7	G/C
chr1	33330343	A	G	FNDC5	58	0.017	AM445	AVASTM	melanoma	7	A/G

Table 4.3 **Patient-specific mutations (truncated)**
This table contains patient-specific mutation lists for patients in the INVAR study. The following cohorts are represented: AVASTM (stage II-III melanoma), MELR (stage IV melanoma) and LUCID (stage I-IIIA NSCLC). Mutation positions are given with the hg19 genome build. The table has been truncated to show the data structure; the full table is available upon request.

Study	Cohort description	Patient ID	Timepoint	Dilution factor	Library prep	Library prep input (ng)	Data type	Mutations analysed	Median raw depth	IR	Evaluable	Mutant reads sum
LUCID	NSCLC stage I-IIIA	40	baseline	NA	Thruplex_Tag_Seq	6.982	custom_ca	233	888	26112	PASS	
LUCID	NSCLC stage I-IIIA	41	baseline	NA	Thruplex_Tag_Seq	8.778	custom_ca	662	1265	102537	PASS	
LUCID	NSCLC stage I-IIIA	42	baseline	NA	Thruplex_Tag_Seq	12.992	custom_ca	203	1596	37684	PASS	1
LUCID	NSCLC stage I-IIIA	43	baseline	NA	Thruplex_Tag_Seq	28.022	custom_ca	196	950	33139	PASS	
LUCID	NSCLC stage I-IIIA	44	baseline	NA	Thruplex_Tag_Seq	5.06	custom_ca	1350	1124	105852	PASS	
LUCID	NSCLC stage I-IIIA	46	baseline	NA	Thruplex_Tag_Seq	11.026	custom_ca	941	563	106835	PASS	
LUCID	NSCLC stage I-IIIA	48	baseline	NA	Thruplex_Tag_Seq	15.334	custom_ca	496	802	72374	PASS	
LUCID	NSCLC stage I-IIIA	49	baseline	NA	Thruplex_Tag_Seq	22.666	custom_ca	386	1041	73872	PASS	
LUCID	NSCLC stage I-IIIA	54	baseline	NA	Thruplex_Tag_Seq	5.402	custom_ca	681	956	76080	PASS	
LUCID	NSCLC stage I-IIIA	55	baseline	NA	Thruplex_Tag_Seq	15.616	custom_ca	330	903	43390	PASS	
LUCID	NSCLC stage I-IIIA	56	baseline	NA	Thruplex_Tag_Seq	13.334	custom_ca	452	1292	90614	PASS	
LUCID	NSCLC stage I-IIIA	39	baseline	NA	Thruplex_Plasma_se	6.616	custom_ca	603	1399	120620	PASS	
LUCID	NSCLC stage I-IIIA	45	baseline	NA	Thruplex_Plasma_se	5.146	custom_ca	127	1425	16110	FAIL	
LUCID	NSCLC stage I-IIIA	47	baseline	NA	Thruplex_Plasma_se	12.018	custom_ca	194	920	22011	PASS	
LUCID	NSCLC stage I-IIIA	50	baseline	NA	Thruplex_Plasma_se	7.962	custom_ca	2349	2003	263573	PASS	
LUCID	NSCLC stage I-IIIA	51	baseline	NA	Thruplex_Plasma_se	11.282	custom_ca	186	2572	36375	PASS	
LUCID	NSCLC stage I-IIIA	52	baseline	NA	Thruplex_Plasma_se	15.436	custom_ca	256	1394	55559	PASS	
LUCID	NSCLC stage I-IIIA	53	baseline	NA	Thruplex_Plasma_se	5.538	custom_ca	292	2014	39345	PASS	

Table 4.4 **Details of libraries sequenced (truncated)**

This table lists all the patient plasma DNA sequencing datasets analysed by INVAR. For each sample/analy details, QC metrics following sequencing, and ctDNA IMAF values following the application of the INVAR non-evaluable if they had no ctDNA signal and <20,000 IR due to the limited sensitivity achieved (Metho truncated to show the data structure; the full table is available upon request. IR, informative reads (see I ratio; IMAF, integrated mutant allele fraction; INVAR score, output of INVAR algorithm based on evidenc equivalent to specificity.

cohort	data type	Min family size	Analysis setting	INVAR score threshold
Stage II-III melanoma	custom_capture	2	Merged reads from 3M and 6M time points where available	0.93945578
Stage II-III melanoma	custom_capture	2	Individual samples at 3M and 6M	0.95492424
Stage I-IIIA NSCLC	custom_capture	2	Individual samples	0.96540881
Stage IV melanoma	custom_capture	2	Individual samples in replicate	0.92043011
Stage IV melanoma	sWGS	1	Individual samples	0.99899498
Stage IV melanoma and stage I-IIIA NSCLC	WES	2	Individual samples	1

Table 4.5 **INVAR score thresholds**
This table gives details on each of the cohorts, the experimental method performed to generate data, and the INVAR score threshold used (determined by ROC analysis, Method 4.7.9). Minimum family size setting was used in read-collapsing (Method 4.7.2).

patient id	days	volume ml	n lesions	Timepoint	evaluable
58	0	23.8895075	1	1	TRUE
58	63	11.3664653	1	3	TRUE
59	0	4.61105083	5	1	TRUE
59	75	1.02110705	5	3	TRUE
59	128	2.17671333	3		TRUE
59	187	1.29680599	5	7	TRUE
59	243	2.60317414	5		TRUE
59	300	5.49328691	5	11	TRUE
59	406	28.2797516	7	16	TRUE
59	483	75.5684028	7		TRUE
59	555	95.386976	7		TRUE
59	609	157.485502	7	18	TRUE
59	720	307.576804	5	22	TRUE
62	0	991.038424	4	1	TRUE
62	57	135.873928	4	3	TRUE
62	114	120.05901	4	6	TRUE
62	169	186.258898	4	7	TRUE
62	224	499.185983	4	9	TRUE
63	0	0.95626485	1	1	FALSE
63	67	0.46475679	1	3	FALSE
63	123	0.49786562	1	5	FALSE
63	176	0.84173131	1	7	FALSE
63	234	1.05460095	1	10	FALSE
65	0	34.916395	5	1	TRUE
65	52	19.8096596	5	2	TRUE
65	106	21.3700037	4	4	TRUE
65	166	46.8142685	3		TRUE
65	189	67.8356401	5	7	TRUE
66	2	70.2068384	3	1	TRUE
66	75	123.342982	4	3	TRUE
66	210	84.3236659	4		TRUE
66	275	56.7803793	3	7	TRUE
66	435	54.7905865	3	11	TRUE
66	487	69.9368477	3	13	TRUE

Table 4.6 **Tumour volume data**
This table shows the CT imaging data for stage IV melanoma patients, processed according to Methods 4.6.10.

Stage	# of patients	Low IR (<20K)	ctDNA detected	ctDNA not detected	Sensitivity	Median IMAF (range)
IA	5	0	2	3	40% (2/5)	$3.95x10^{-04}$ ($2.3x10^{-4}$-$5.6x10^{-4}$)
IB	6	1	3	2	60% (3/5)	$3.27x10^{-04}$ ($2.47x10^{-4}$-$3.43x10^{-3}$)
II	6	1	5	0	100% (5/5)	$2.46x10^{-03}$ ($3.66x10^{-4}$-$1.94x10^{-2}$)
IIIA	2	0	2	0	100% (2/2)	$4.99x10^{-03}$ ($1.27x10^{-4}$-$9.85x10^{-3}$)
Total	19	2	12	5	70.5% (12/17)	$6.79x10^{-4}$ ($1.27x10^{-4}$-$1.94x10^{-2}$)

Table 4.7 **Detection summary for stage I-IIIA NSCLC cohort**
Detection and median IMAF values are shown by stage for the stage I-IIIA NSCLC cohort. Totals for the whole NSCLC cohort are shown in the bottom row. IR, informative reads; IMAF, integrated mutant allele fraction.

	Total	
	N	%
Characteristics		
	38	
Sex		
Male	17	45
Female	21	55
Breslow thickness		
<=2.0mm	16	42
>2-4.0mm	8	21
>4.0mm	12	32
Unknown	2	5
Ulceration		
Present	9	24
Absent	23	60
Unknown	6	16
Disease stage		
II	8	21
IIIA	4	10
IIIB	17	45
IIIC	9	24
N classification		
II (No or N/A)	8	21
III (N1a and N2a)	6	16
III (other N)	24	63
ECOG performance status		
0	35	92
1	3	8
BRAF and NRAS mutation		
BRAF mutation	19	50
NRAS mutation	8	21
BRAF and NRAS WT	10	26
BRAF WT and NRAS not tested	1	3
Trial arm		
Observation	38	100
Time from latest surgery to trial entry in weeks		
Median (Range)	11 (4-12)	

Table 4.8 **Summary of patient characteristics in the stage II-III melanoma cohort** Patient characteristics are shown for the stage II-III melanoma cohort (n = 38).

Chapter 5

Discussion

5.1 Overview

In this thesis, I describe an approach I developed in collaboration with my colleagues (listed in the attribution sections throughout) for enhanced detection and monitoring of ctDNA. Given a patient-specific mutation list, INVAR aggregates signal across a large number of loci, and weights signal based on characteristics of patient-specific ctDNA sequencing data. We demonstrate that this method is not limited to custom capture data, but may also be applied to WES and WGS data. The INVAR algorithm (Chapter 4) builds upon our understanding of mutation representation in plasma and patient-specific sequencing generated by amplicon sequencing (Chapter 3).

This thesis focuses on patient-specific sequencing data, where between 10^1-10^4 known tumour-mutated loci were interrogated using plasma sequencing. When interrogating multiple mutations, signal may be considered either at each individual locus, or signal may be aggregated across all loci (taking into account the differences in error rate and other characteristics between loci). Initially, amplicon sequencing (TAm-Seq) was used to target up to 62 mutations per patient, then we moved towards using broader

sequencing methods (i.e. custom-capture sequencing, WES and WGS) in order to generate plasma sequencing data at up to 5,036 patient-specific mutations. At the same time as moving to capture sequencing, in order to reduce background error rates, we used molecularly barcoded library preparation kits (whole-genome amplification kits). The differences between the sequencing data are shown in Table 5.1.

By combining both error-suppression and a large number of mutations with a novel analytical method using hybrid-capture data, we improved sensitivity for ctDNA from conventional levels of 0.01% mutant allele fraction to 0.0001% (1 ppm). This approach utilised the learning from our TAm-Seq work which showed the benefit of targeting multiple mutations, and that private mutations are represented less well in the plasma. Moving to non-amplicon-based sequencing methods allowed exploratory analyses into the relationship between fragment size and mutation detection, which has previously been studied [198, 65]. Together, these learning points were combined and formalised as the INtegration of VAriant Reads algorithm, a method for sensitive detection and quantification of ctDNA in a sample using a list of patient-specific mutations. Aside from demonstrating 1-2 orders of magnitude greater sensitivity in plasma than previous studies [106], we also showed that by targeting a large number of loci, samples with limited input DNA (and limited sequencing, i.e. sWGS) can be leveraged for sensitive detection to 0.01%, again orders of magnitude of more sensitive than copy-number-based analyses of sWGS data [115].

5.2 Patient-specific amplicon sequencing

Patient-specific amplicon sequencing panels for 9 patients with stage IV melanoma were designed based on tumour exome sequencing of metastatic tumour biopsies. Suzanne Murphy and I performed amplicon sequencing on longitudinal plasma samples

	Amplicon sequencing	Hybrid-capture sequencing	WGS
Number of mutations	10-100	10^2-10^4	>10^2-10^4
Error-suppression	No	Yes	Yes*
Fragment size information	No	Yes	Yes
Samples per lane of HiSeq4000 (to achieve ∼1,000x depth)	>300	∼5-10†	1–30 (0.5-15x depth)

Table 5.1 **Properties of sequencing data generated**
In this thesis, amplicon, hybrid-capture and whole-genome sequencing data were generated and analysed with patient-specific detection algorithms.
*Error-suppression by read-collapsing can only be performed with WGS data if the library input mass is restricted, thereby allowing families of sufficient size to be generated.
†Dependent on off-target rate of the capture kit.

from those patients and re-sequenced the tumour DNA eluates with the same method to confirm the allele fractions of those mutations. These data allowed two separate analyses: in Section 3.5.1, we performed comparisons of mutation profiles between matched plasma and tumour samples in order to assess mutation representation in plasma; and in Section 3.5.2, we analysed the longitudinal ctDNA profiles for these patients, assessing the sensitivity of amplicon sequencing with individual and multiple mutations.

5.2.1 Tumour vs. plasma comparison

In Chapter 3, Amplicon sequencing of matched tumour and plasma samples allowed comparisons between mutation profiles with >10,000x depth per locus. This analysis highlighted that while longitudinal metastatic tumour biopsies showed variable mutation profiles, mutation profiles of longitudinal plasma samples from the same patients showed a lesser degree of change over time. Of course, these samples are temporally separated and are thus subject to clonal evolution. However, sampling error

may confound such analyses. Therefore, we suggest that combined analysis of tumour samples and matched plasma samples would be most informative in order to interpret tumour mutation profile data. One possible future approach could be to correct tumour allele fractions by plasma allele fractions, i.e. weighting tumour mutation profiles by their representation in plasma.

However, this analysis is currently limited, at present, by not knowing the ground truth as to whether: (a) the plasma mutation profiles are truly representative of the tumour, and the tumour biopsy is subject to sampling error, or (b) the tumour biopsies are accurate, but the factors affecting release of ctDNA vary such that mutation profiles in plasma remain relatively constant between samples. In future, comparing each of the mutation profiles against FFPE tumour mutation profiles would provide an additional sample against which to compare. Alternatively, instead of trying to show which sample type is more 'true', we can accept that both are confounded each in their own ways. Thus, it might be appropriate to generate mutation profiles that consist of the product the mutation profiles of the tumour and plasma. This approach would effectively assign equal weight to tumour and plasma samples, and might overcome some of the sampling biases from each, allowing for more accurate monitoring of clonal evolution. The mutation detection approach in each of the sample types would still remain separate, in case tumour samples and cfDNA samples show different error profiles.

5.2.2 Sensitivity analysis from multiple mutations

After comparing time points where we had matched tumour and plasma samples, we then analysed longitudinal plasma samples using the same method of patient-specific TAm-Seq. In this analysis, we showed preliminary evidence for the sensitivity benefit of targeting multiple mutations to overcome sampling error, using both a dilution

series of patient plasma cfDNA (Fig. 3.4) and clinical samples (Fig. 3.7). By targeting multiple patient-specific loci, we detected and quantified ctDNA to a mean mutant allele fraction of 0.01%, over an order of magnitude lower than what was shown in the original TAm-Seq study [29]. This level of sensitivity is approaching the theoretical limit of detection given the background error rates of this method, indicating that to improve detection rates further, background error rates must be suppressed.

We did not attempt to improve background error rates of TAm-Seq in this study. This could have been performed experimentally, for example, by adapting the experimental protocol to include molecular barcodes through ligation prior to the rest of the TAm-Seq workflow (the published depicted in Fig. 5.1). Alternatively, error-suppression to a lesser degree could be achieved (with the existing data) through generating the consensus sequence between the forward and reverse mates from a paired-end read, with a tool such as 'Paired-End reAd mergeR' [286], as was applied to the Abbosh et al. amplicon sequencing study [117]. Paired-end read merging is particularly suited to amplicon sequencing of cfDNA data because of the short fragment size of cfDNA, combined with the even smaller amplicon size (~90bp), since any fragment shorter than 150bp would be read entirely in both directions (using PE150 sequencing). Despite not implementing PEAR in any of the methods in this thesis, the concept of error-suppression based on F and R mates was later used in the INVAR algorithm since any mutant signal was required to be present in both reads of the pair (Method 4.7.3).

To improve the sensitivity of TAm-Seq would require both molecular barcodes and a larger number of mutations to be targeted. However, the number of mutations targetable with TAm-Seq is limited since during the pre-amplification step, if too many primer pairs are multiplexed, they may form primer-dimers, or show reduced efficiency of amplification [288].

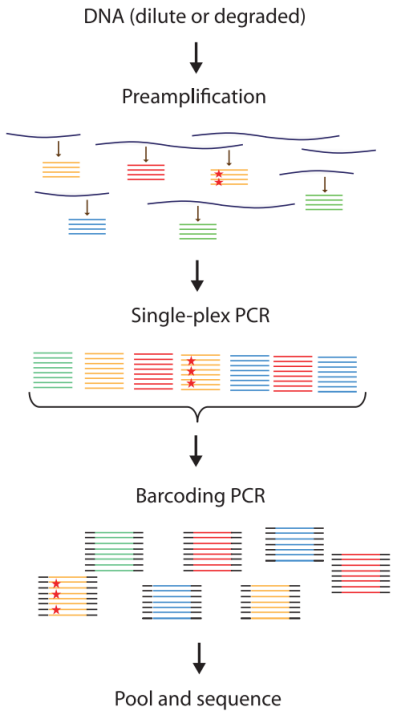


Fig. 5.1 **TAm-Seq outline**
Dilute or degraded cfDNA is pre-amplified by PCR to amplify rare events. Each amplicon of interest then undergoes a single-plex PCR in order to achieve similar coverage across PCR amplicons. Next, a barcoding PCR is carried out, where sample barcodes and Illumina adaptors are added. Molecular barcodes would need to be annealed to each individual molecule, then a limited PCR amplification is performed prior to the pre-amplification step. Taken from [29].

5.3 Patient-specific capture sequencing

In order to target a larger number of loci than was possible with TAm-Seq, we next considered patient-specific hybrid-capture approaches, which can target megabases of the genome. Balancing of bait concentrations is handled by the manufacturer, so that validation of individual baits does not have to be performed again in our lab. In addition, we could leverage existing molecularly barcoded library preparation kits to reduce error rates. Lastly, capture sequencing data also retains greater information on the starting fragment, such as the fragment size and its start and end position.

Starting with the same patient-specific mutations lists generated for the 9 stage IV melanoma patients in Chapter 3, we generated a custom sequencing panel where all mutations passing relaxed QC filters were included (Method 4.7.1). We aimed to maximise the number of mutations targeted, since we initially wanted to explore the factors that influence the representation of a given mutation in plasma. This relaxed filtering strategy still persists in our design of custom capture panels, since we can now assign weight to mutation signals based on the confidence of the tumour call rather than restricting the mutation list and losing the signal permanently.

We designed custom sequencing panels using Agilent SureDesign software and Agilent SureSelectXT capture kits (Method 4.6.6). We generated sequencing data with 1,000-2,000x depth at each of the patient-specific loci of interest; these depth values will undoubtably rise as sequencers with greater output become available, and as kits with lower off-target rates are used. Eventually, if sequencing becomes cheap enough, WES or WGS approaches would allow the panel design step to be avoided altogether, as we suggest in Fig. 4.12. However, saturating levels of sequencing for WGS are currently too costly, and so patient-specific panels will give deeper sequencing at loci of interest for the foreseeable future.

5.3.1 Overcoming limited data for background-error estimation

One of the challenges of patients-specific sequencing data is that of generating adequate control data for every patient-specific locus. For fixed sequencing panels, the same panel can be run on many samples to accurately characterise differences between mutations, though this process must be repeated for each new patient-specific panel. Therefore, we opted to assess background error rates not only per locus, but also per trinucleotide context, based on findings that each mutation class has its own error rate [106]. This allowed identification of the noisiest loci (based on locus-specific

noise), and for loci with zero background error in control samples, we leveraged the background error rate of that context. Of course, greater granularity in error rates could be obtained by assessment of contexts of >3nt (e.g. 5nt context), though the compute required increases rapidly with each additional base pair in the context.

5.3.2 Reduction of background noise further

Aside from using molecular barcoding for error-suppression, we developed additional filtering steps to reduce error further. Firstly, We required signal to be present in both F and R mates of the paired end read. This served the dual purpose of reducing background error rates (Fig. 4.5), and also excludes long cfDNA fragments from the analysis (since fragments >300bp simply cannot have any overlapping reads using PE150 sequencing). This effectively acts as a size-selection step, which may enhance detection of the early-stage cancer samples, since their mutant signal arose entirely from fragments less than 200bp. In advanced disease, we observed mutant signal in fragments of length indicative of dinucleosomal cfDNA; these fragments would be lost, but ctDNA levels were sufficiently high such that detection was still supported by short fragments.

In addition, we also developed a method of outlier-suppression for loci that showed signal greatly above the rest of the patient-specific loci in that sample (Method 4.7.5). This was based on the observation in our data that at the lowest levels of ctDNA, sampling error is expected, and that signal should only ever look like individual molecules at individual loci, and should not be multiple mutant reads all at one locus when all other loci are negative. This approach is only possible when a patient-specific mutation list of 100 to >1000 mutations is available, because it relies on identifying loci that are different in allele fraction to the remainder. We find that in our data, noise is reduced to ~50% while >90% of signal in patient samples is retained.

5.3.3 Signal-weighting

In order to detect the lowest levels of ctDNA, it is important to maximise the number of informative reads (IR, collapsed reads overlying patient-specific mutation loci, Fig. 4.4), since the limit of detection achievable is related to the reciprocal of IR. Thus, instead of using hard thresholds on fragment size or tumour allele fraction, which would result in increases in observed ctDNA signal (but could cause loss of the signal entirely, as shown by Fan et al. [208]), we instead opted for an approach of weighting signal by tumour allele fraction (Fig 4.8) and fragment size (Fig. 4.7).

Tumour loci were weighted based on tumour allele fraction, which varies between 0 and 1, and thus provides a straightforward approach for weighting based on the product of the signal and the tumour allele fraction. However, it may not always be the case that a higher allele fraction mutation should be assigned greater confidence than those with lower allele fraction, since the number of mutant reads supporting the call, the mutation context etc. should also be considered. Thus, in future, it may be more appropriate to use a probability given by the mutation caller to perform weighting. For example, Mutect2 outputs a 'LOD score', which is a log-odds score for the confidence in the mutation based on the number of mutant reads given the depth at that locus and the error rate [289]. However, in order to utilise LOD scores in the current INVAR code, each value would have to be translated to values between 0 and 1, as LOD scores can range between <10 to 1 x 10^6 [290].

Signal was further enhanced based on fragment size profiles, since it is known that short fragments are more likely to be tumour-derived [65]. Previous analyses have used size-selection approaches [199], though hard cut-offs for signal may result in true mutant signal being lost [208]. Therefore, we opted for a weighting-based approach in order to retain such signal by assigning weight based on what we know about the

enrichment ratio for ctDNA for a given fragment size. However, the current challenges of this weighting-based approach are as follows:

- Firstly, size-weights are assigned based on all the size data available for that cancer type and stage. Although the size profile is smoothed to compensate for limited data (Method 4.7.7), any deviation from the ground truth confounds the analysis.
- Currently, size-weighting is performed per sample relative to the fragment size profile of that cohort (Method 4.7.6). Although the size profile is smoothed, there is still some data in the profile that is from that same sample, meaning that the size-weighting is partially circular. This should be addressed with a leave-one-out style of analysis whereby the sample is weighted based on the size profiles of all other samples only. This would take longer to compute since the fragment size profile of the $n-1$ cohort would be recalculated each time.

5.3.4 Biological limitations of sensitive detection

Based on the INVAR method of aggregation and weighting of signal, we were able to quantify ctDNA to parts per million (Fig. 4.13) in early- and advanced-stage disease. Quantification to parts per million was made possible by sequencing millions of informative reads (IR), which was only possible by knowing tumour mutations in advance. To boost sensitivity further, the number of IR will have to increase correspondingly. For example, in order to reach 10^7 IR, 10^4 loci would need to be identified (only WGS can achieve this), and 3,000x deduplicated coverage (or haploid genomes analysed, hGA) needs to be reached (Fig. 4.4).

We suggest that 10^7-10^8 IR is the upper limit of the sensitivity of INVAR. The error rate is not the limiting factor, since background error rates down to 10^{-12} can be reached

through duplex sequencing [99, 106]. Instead, the expected number of mutations per whole genome of a cancer [234] and the volume of blood possible to draw at once becomes limiting. In this study, melanoma and NSCLC patient samples were used in order to maximise the number of detected mutations per exome, and thereby maximise the sensitivity achieved with this approach. To achieve adequate sensitivity in cancer types such as renal cancer or ovarian (which have low mutation rates) would require tumour WGS to be performed since tumour WES would detect an insufficient number of mutations to reach the levels of sensitivity described in this thesis. Furthermore, the maximum blood volume that it is feasible to take at once in clinic (<30-50mL) places a hard limit on sensitivity. Larger volumes may be taken using methods such as plasmapheresis, though this may be impractical and difficult to scale for routine use. Alternatively, larger volumes may be taken if patients were to have blood sampling on consecutive days, but again this may not be practical or acceptable for patients.

It is possible that we are approaching the maximum possible sensitivity of ctDNA-based analyses, and that further improvements would require a multi-modal approach whereby other nucleic acids or proteins are targeted in addition to ctDNA. This has been borne out by recent developments, which have shown that platelet RNA can be used to accurately classify both early and advanced stage cancer patients [221]. In addition, a combined protein and ctDNA-based approach showed improved detection relative to ctDNA alone [291]. The ideal combination of tests would utilise every part of the blood sample with minimal waste, thereby maximising the sensitivity achieved from the fewest blood tubes possible.

5.3.5 Application of INVAR to limited-input samples

By targeting a large number of mutations, INVAR can aggregate signal across loci when there are few hGA. Low hGA can arise from low input DNA, or low sequencing

depth, or both, since the number of error-suppressed families generated is a function of input DNA mass and the extent of oversequencing. We show that ctDNA can be quantified to 10^{-4} from sWGS data (equivalent to 0.6 hGA), and similarly detect ctDNA to 10^{-4} in custom capture sequencing data that was downsampled to 10hGA. Since 10hGA may be obtained from the volume of plasma from a drop of blood, we suggest that INVAR might enable monitoring using limited blood volumes. This would allow pin-prick diagnostics, and potentially could allow self-collection of blood by patients onto card, which may be later extracted in a centralised laboratory.

So far, we have shown that given >0.6 hGA, the INVAR algorithm can quantify disease to 10^{-4}. However, we have not yet proven whether >0.6 hGA can be obtained from a blood drop. Theoretically, one blood drop may yield up to 60 genome copies, assuming a droplet volume of 75µL, plasma content of 55%, and cfDNA concentration of 1-5ng/mL [279]; with a library preparation efficiency of 30%, this may yield 4-20 hGA. Thus, to address this, we will perform DNA extraction from dried blood spots and will apply INVAR. We will start with dried blood spots, rather than miniscule volumes of blood, since blood spots can be easily stored on Whatman FTA (filter paper) cards at ambient temperature for at least 11 years [292], and can be easily shipped. Dealing with miniscule blood volumes would require the use of specialised collection devices [293], which may enable multiple potential applications, but are less readily available than filter paper cards, and would cost more to manufacture and ship.

For these future blood spot experiments, I expect that there will be dilution of signal caused by white blood cell lysis and release of genomic DNA (gDNA). Thus, we will start by analysis of high ctDNA samples in order to quantify any potential drop in ctDNA IMAF due to this dilution effect. At the same time, we will characterise the fragment size profile of these samples to try identify any potential differences in size profile between cfDNA and fragmented gDNA, which could indicate whether an

in vitro or *in silico* size-selection are warranted, in order to exclude as many gDNA fragments as possible.

5.4 Conclusion

INVAR can be applied to reach sensitivity to parts per million, or can allow monitoring with limited input amounts down to the levels of sensitivity that are commonly applied today by other methods. INVAR depends on the availability of previous mutation lists from the patient's cancer sequencing (demonstrated here from tumour sequencing, but this can be obtained at higher costs from plasma sequencing when ctDNA levels are high). Thus, this approach addresses a specific clinical context and need, i.e. for sensitive monitoring following surgery or initation of immunotherapy, where tumour mutation information is available.

Patients increasingly undergo broader and deeper tumour sequencing, either to identify actionable mutations [294, 295], to assess tumour mutation burden for immunotherapies [294], or in large clinical or genomic studies where WGS is being performed [296]. We suggest that as sequencing costs decline further, INVAR-like algorithms could have increasing applicability and utility for sensitive monitoring either from conventional or limited blood input volumes.

Chapter 6

Publications

6.1 Published works

6.1.1 Manuscripts

Wan, JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, Pacey S, Baird R & Rosenfeld N (2017). Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nature Reviews Cancer*, 17(4), 223-238.

Mouliere F, Piskorz A, Chandrananda D, Moore E, Morris J, Smith CG, Goranova T, Heider K, Mair R, Supernat A, Gounaris I, Ros S, **Wan JCM**, Jimenez-Linan M, Gale D, Brindle K, Massie CE, Parkinson CA, Brenton JD, Rosenfeld N (2017). Selecting Short DNA Fragments In Plasma Improves Detection Of Circulating Tumour DNA *bioRxiv* 134437; [Available online: https://doi.org/10.1101/134437]

6.1.2 Abstracts

Wan JCM, Murphy S, Gale D, Morris JA, Mouliere F, Gill A, Marass F, Heider K, Parkinson C, Gallagher F, Durrani A, McDermott U, Massie CE, Corrie P, and Rosenfeld N (2017). *Journal of Clinical Oncology* 35:15_suppl, e21032-e21032 [Available online: http://ascopubs.org/doi/abs/10.1200/JCO.2017.35.15_suppl.e21032]

Wan JCM, Murphy S, Gale D, Morris JA, Mouliere F, Gill A, Marass F, Heider K, Chandrananda D, Parkinson C, Gallagher F, Durrani A, McDermott U, Massie CE, Corrie P, and Rosenfeld N (2017). *The Lancet* 39:S99 [Available online: http://dx.doi.org/10.1016/S0140-6736(17)30495-6]

6.1.3 Patents

Wan JCM, Heider K, Fisher E, Mouliere F, Smith C, Massie CE, Rosenfeld N (2018). Integration of Variant Reads. Filed March 6th 2018.

6.2 Unpublished works

Wan JCM*, Heider K*, Gale D, Murphy S, Fisher E, Ruiz-Valdepenas A, Morris J, Mouliere F, Chandrananda D, Marshall A, Gill AB, Chan P-Y, Qian W, Doughton G, Barker E, Young G, Cooper W, Hudecova I, Marass F, Bignell GR, Alifrangis C, Eisen T, Middleton MR, Gallagher FA, Parkinson C, Durrani A, McDermott U, Rassl DM, Harden S, Smith CG, Rintoul RC, Massie†, Corrie PG†, Rosenfeld N†,§ (Re-submitted in July 2018).

Monitoring ctDNA in low burden cancer to parts per million by integration of variant reads across thousands of mutated loci

*Equally contributing authors

†CM, PGC and NR jointly supervised this

§Corresponding author

Wan JCM, Murphy S, Gale D, Morris J, Mouliere F, Marass F, Heider K, Chandrananda D, Smith CG, Bignell GR, Alifrangis CA, Parkinson C, Durrani A, McDermott U, Massie C, Corrie PG, Rosenfeld N§ (Submitted in November 2016).
Plasma DNA analysis mitigates biased mutation detection in tumour biopsies
§Corresponding author

References

[1] P Mandel and P Métais. Les acides nucléiques du plasma sanguin chez l'homme. *Comptes rendus des seances de la Societe de biologie et de ses filiales*, 142(3-4):241–243, 2 1948.

[2] S. A. Leon, B. Shapiro, D. M. Sklaroff, and M. J. Yaros. Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Research*, 37(3):646–650, 1977.

[3] Florent Mouliere, Safia El Messaoudi, Dalong Pang, Anatoly Dritschilo, and Alain R. Thierry. Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. *Molecular Oncology*, 8(5):927–941, 2014.

[4] Edison Moraes Rodrigues Filho, Daniel Simon, Nilo Ikuta, Caroline Klovan, Fernando Augusto Dannebrock, Carla Oliveira de Oliveira, and Andrea Regner. Elevated Cell-Free Plasma DNA Level as an Independent Predictor of Mortality in Patients with Severe Traumatic Brain Injury. *Journal of Neurotrauma*, 31(19):1639–1646, 2014.

[5] Nai-Wen Tsai, Tsu-Kung Lin, Shang-Der Chen, Wen-Neng Chang, Hung-Chen Wang, Tzu-Ming Yang, Yu-Jun Lin, Chung-Ren Jan, Chi-Ren Huang, Chia-Wei Liou, and Cheng-Hsien Lu. The value of serial plasma nuclear and mitochondrial DNA levels in patients with acute ischemic stroke. *Clinica Chimica Acta*, 412(5-6):476–479, 2011.

[6] Sarah Breitbach, Suzan Tug, Susanne Helmig, Daniela Zahn, Thomas Kubiak, Matthias Michal, Tommaso Gori, Tobias Ehlert, Thomas Beiter, and Perikles Simon. Direct Quantification of Cell-Free, Circulating DNA from Unpurified Plasma. *PLoS ONE*, 9(3):e87838, 2014.

[7] Iwijn De Vlaminck, Hannah A Valantine, Thomas M Snyder, Calvin Strehl, Garrett Cohen, Helen Luikart, Norma F Neff, Jennifer Okamoto, Daniel Bernstein, Dana Weisshaar, Stephen R Quake, and Kiran K Khush. Circulating cell-free DNA enables noninvasive diagnosis of heart transplant rejection. *Science translational medicine*, 6(241):241ra77, 2014.

[8] Iwijn De Vlaminck, Lance Martin, Michael Kertesz, Kapil Patel, Mark Kowarsky, Calvin Strehl, Garrett Cohen, Helen Luikart, Norma F Neff, Jennifer Okamoto, Mark R Nicolls, David Cornfield, David Weill, Hannah Valantine,

Kiran K Khush, and Stephen R Quake. Noninvasive monitoring of infection and rejection after lung transplantation. *Proceedings of the National Academy of Sciences*, 112(43):13336–41, 2015.

[9] Y. M. D. Lo, Noemi Corbetta, Paul F. Chamberlain, Vik Rai, Ian L. Sargent, Christopher W G Redman, and James S. Wainscoat. Presence of fetal DNA in maternal plasma and serum. *Lancet*, 350(9076):485–487, 1997.

[10] Jon A. Hyett, Glenn Gardener, Taita Stojilkovic-Mikic, Kirstin M. Finning, Peter G. Martin, Charles H. Rodeck, and Lyn S. Chitty. Reduction in diagnostic and therapeutic interventions by non-invasive determination of fetal sex in early pregnancy. *Prenatal Diagnosis*, 25(12):1111–1116, 2005.

[11] H Saito, A Sekizawa, T Morimoto, M Suzuki, and T Yanaihara. Prenatal DNA diagnosis of a single-gene disorder from maternal plasma. *Lancet*, 356(9236):1170, 2000.

[12] Y M Dennis Lo, Fiona M F Lun, K C Allen Chan, Nancy B Y Tsui, Ka C Chong, Tze K Lau, Tak Y Leung, Benny C Y Zee, Charles R Cantor, and Rossa W K Chiu. Digital PCR for the molecular detection of fetal chromosomal aneuploidy. *Proceedings of the National Academy of Sciences*, 104(32):13116–13121, 2007.

[13] Megan Allyse, Mollie A. Minear, Elisa Berson, Shilpa Sridhar, Margaret Rote, Anthony Hung, and Subhashini Chandrasekharan. Non-invasive prenatal testing: A review of international implementation and challenges. *International Journal of Women's Health*, 7:113–126, 2015.

[14] Melissa Hill, David Wright, Rebecca Daley, Celine Lewis, Fiona McKay, Sarah Mason, Nicholas Lench, Abigail Howarth, Christopher Boustred, Kitty Lo, Vincent Plagnol, Kevin Spencer, Jane Fisher, Mark Kroese, Stephen Morris, and Lyn S Chitty. Evaluation of non-invasive prenatal testing (NIPT) for aneuploidy in an NHS setting: a reliable accurate prenatal non-invasive diagnosis (RAPID) protocol. *BMC Pregnancy and Childbirth*, 14(1):229, 2014.

[15] M. Stroun, P. Anker, P. Maurice, J. Lyautey, C. Lederrey, and M. Beljanski. Neoplastic Characteristics of the DNA Found in the Plasma of Cancer Patients. *Oncology*, 46(5):318–322, 1989.

[16] David Sidransky, Takashi Tokino, Stanley R Hamilton, Kenneth W Kinzler, Bernard Levin, Philip Frost, and Bert Vogelstein. Identification of ras oncogene mutations in the stool of patients with curable colorectal tumors. *Science (New York, N.Y.)*, 256(5053):102–5, 1992.

[17] Carlos Caldas, Stephan A Hahn, Ralph H Hruban, A Hahn, Mark S Redston, J Yeo, and Scott E Kern. Detection of K-ras mutations in the stool of patients with pancreatic adenocarcinoma and pancreatic ductal hyperplasia. *Cancer Research*, 54(13):3568–3573, 1994.

[18] Li Mao, Ralph H Hruban, Jay O Boyle, Melvyn Tockman, and David Sidransky. Detection of Oncogene Mutations in Sputum Precedes Diagnosis of Lung Cancer. *Cancer Research*, 54(7):1634–1637, 1994.

[19] Satoshi Takeda, Shigetoshi Ichii, and Yusuke Nakamura. Detection of K-ras mutation in sputum by mutant-allele-specific amplification (MASA). *Human Mutation*, 2(2):112–117, 1 1993.

[20] G D Sorenson, D M Pribish, F H Valone, V A Memoli, D J Bzik, and S L Yao. Soluble normal and mutated DNA sequences from single-copy genes in human blood. *Cancer Epidemiology Biomarkers & Prevention*, 3(1):67 LP – 71, 1 1994.

[21] Elizabeth M. Swisher, Melissa Wollan, Sarita M. Mahtani, Julia B. Willner, Rochelle Garcia, Barbara A. Goff, and Mary Claire King. Tumor-specific p53 sequences in blood and peritoneal fluid of women with epithelial ovarian cancer. *American Journal of Obstetrics and Gynecology*, 193(3):662–667, 2005.

[22] Hideharu Kimura, Kazuo Kasahara, Makoto Kawaishi, Hideo Kunitoh, Tomohide Tamura, Brian Holloway, and Kazuto Nishio. Detection of Epidermal Growth Factor Receptor Mutations in Serum as a Predictor of the Response to Gefitinib in Patients with Non-Small-Cell Lung Cancer. *Clin Cancer Res*, 12(13):3915–3921, 2006.

[23] Gabriella Sozzi, Katia Musso, Cathy Ratcliffe, Peter Goldstraw, and Marco A Pierotti. Detection of Microsatellite Alterations in Plasma DNA of Non-Small Cell Lung Cancer Patients: A Prospect for Early Diagnosis. *Clin Cancer Res*, 5(10):2689–2692, 1999.

[24] Thierry Lecomte, Anne Berger, Franck Zinzindohoué, Stéphanie Micard, Bruno Landi, Hélène Blons, Philippe Beaune, Paul Henri Cugnenc, and Pierre Laurent-Puig. Detection of free-circulating tumor-associated DNA in plasma of colorectal cancer patients and its association with prognosis. *International Journal of Cancer*, 100(5):542–548, 2002.

[25] B Vogelstein and K W Kinzler. Digital PCR. *Proceedings of the National Academy of Sciences*, 96(16):9236–9241, 1999.

[26] Devin Dressman, Hai Yan, Giovanni Traverso, Kenneth W Kinzler, and Bert Vogelstein. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proceedings of the National Academy of Sciences of the United States of America*, 100(15):8817–8822, 2003.

[27] Frank Diehl, Meng Li, Devin Dressman, Yiping He, Dong Shen, Steve Szabo, Luis a Diaz, Steven N Goodman, Kerstin a David, Hartmut Juhl, Kenneth W Kinzler, and Bert Vogelstein. Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proceedings of the National Academy of Sciences*, 102(45):16368–16373, 2005.

[28] Frank Diehl, Kerstin Schmidt, Michael a Choti, Katharine Romans, Meng Li, Katherine Thornton, Nishant Agrawal, Lori Sokoll, A Steve, Kenneth W Kinzler, Bert Vogelstein, and Luis a Diaz Jr. Circulating mutant DNA to assess tumor dynamics. *Nat. Med.*, 14(9):985–990, 2008.

[29] Tim Forshew, Muhammed Murtaza, C. Parkinson, D. Gale, D. W. Y. Tsui, F. Kaper, S.-J. Sarah-Jane Dawson, A. M. Piskorz, Mercedes Jimenez-Linan, David R Bentley, J. Hadfield, A. P. May, Carlos Caldas, James D. Brenton, and Nitzan Rosenfeld. Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA. *Science Transl. Med.*, 4(136):68–136, 5 2012.

[30] Sarah-Jane Dawson, Dana W Y Tsui, Muhammed Murtaza, Heather Biggs, Oscar M Rueda, Suet-Feung Chin, Mark J Dunning, Davina Gale, Tim Forshew, Betania Mahler-Araujo, Sabrina Rajan, Sean Humphray, Jennifer Becq, David Halsall, Matthew Wallis, David Bentley, Carlos Caldas, and Nitzan Rosenfeld. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *The New England Journal of Medicine*, 368(13):1199–209, 3 2013.

[31] Rebecca J Leary, Mark Sausen, Isaac Kinde, Nickolas Papadopoulos, D John, David Craig, Joyce O Shaughnessy, Kenneth W Kinzler, Bert Vogelstein, Luis a Diaz Jr, and Victor E Velculescu. Detection of Chromosomal Alterations in the Circulation of Cancer Patients with Whole-Genome Sequencing. *Science Transl. Med.*, 4(162):162ra154, 2012.

[32] K C A Chan, Peiyong Jiang, Carol W M Chan, Kun Sun, John Wong, Edwin P Hui, Stephen L Chan, Wing Cheong Chan, David S C Hui, Simon S M Ng, Henry L Y Chan, Cesar S C Wong, Brigette B Y Ma, Anthony T C Chan, Paul B S Lai, Hao Sun, Rossa W K Chiu, and Y M Dennis Lo. Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. *Proceedings of the National Academy of Sciences*, 110(47):18761–8, 2013.

[33] Ellen Heitzer, Martina Auer, Eva Maria Hoffmann, Martin Pichler, Christin Gasch, Peter Ulz, Sigurd Lax, Julie Waldispuehl-Geigl, Oliver Mauermann, Sumitra Mohan, Gunda Pristauz, Carolin Lackner, Gerald Höfler, Florian Eisner, Edgar Petru, Heinz Sill, Hellmut Samonigg, Klaus Pantel, Sabine Riethdorf, Thomas Bauernhofer, Jochen B Geigl, and Michael R Speicher. Establishment of tumor-specific copy number alterations from plasma DNA of patients with cancer. *International Journal of Cancer*, 133(2):346–56, 2013.

[34] Muhammed Murtaza, Sarah-Jane Dawson, Dana W. Y. Tsui, Davina Gale, Tim Forshew, Anna M. Piskorz, Christine Parkinson, Suet-Feung Chin, Zoya Kingsbury, Alvin S. C. Wong, Francesco Marass, Sean Humphray, James Hadfield, David Bentley, Tan Min Chin, James D Brenton, Carlos Caldas, and Nitzan Rosenfeld. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature*, 497(7447):108–12, 5 2013.

[35] A. R. Thierry, S. El Messaoudi, P. B. Gahan, P. Anker, and M. Stroun. Origins, structures, and functions of circulating DNA in oncology. *Cancer and Metastasis Reviews*, 35(3):347–376, 2016.

[36] Basant Kumar Thakur, Haiying Zhang, Annette Becker, Irina Matei, Yujie Huang, Bruno Costa-Silva, Yan Zheng, Ayuko Hoshino, Helene Brazier, Jenny Xiang, Caitlin Williams, Ruth Rodriguez-Barrueco, Jose M Silva, Weijia Zhang,

Stephen Hearn, Olivier Elemento, Navid Paknejad, Katia Manova-Todorova, Karl Welte, Jacqueline Bromberg, Héctor Peinado, and David Lyden. Double-stranded DNA in exosomes: a novel biomarker in cancer detection. *Cell Research*, 24(6):766–9, 2014.

[37] Christoph Kahlert, Sonia A. Melo, Alexei Protopopov, Jiabin Tang, Sahil Seth, Moritz Koch, Jianhua Zhang, Juergen Weitz, Lynda Chin, Andrew Futreal, and Raghu Kalluri. Identification of Double-stranded Genomic DNA Spanning All Chromosomes with Mutated KRAS and p53 DNA in the Serum Exosomes of Patients with Pancreatic Cancer. *Journal of Biological Chemistry*, 289(7):3869–3875, 2014.

[38] Sabine Jahr, Hannes Hentze, Sabine Englisch, Dieter Hardt, Frank O Fackelmayer, Rolf-Dieter Hesch, and Rolf Knippers. DNA Fragments in the Blood Plasma of Cancer Patients: Quantitations and Evidence for Their Origin from Apoptotic and Necrotic Cells. *Cancer Research*, 61:1659–1665, 2001.

[39] M. Stroun, J. Lyautey, C. Lederrey, A. Olson-Sand, and P. Anker. About the possible origin and mechanism of circulating DNA: Apoptosis and active DNA release. *Clinica Chimica Acta*, 313:139–142, 2001.

[40] I. Botezatu, O. Serdyuk, G. Potapova, V. Shelepov, R. Alechina, Y. Molyaka, V. Ananaposev, I. Bazin, A. Garin, M. Narimanov, V. Knysh, H. Melkonyan, S. Umansky, and A. Lichtenstein. Genetic Analysis of DNA Excreted in Urine: A New Approach for Detecting Specific Genomic DNA Sequences from Cells Dying in an Organism. *Clinical Chemistry*, 46(8):1078–1084, 2000.

[41] K. C Allen Chan, Sing F. Leung, Sze W. Yeung, Anthony T C Chan, and Y. M Dennis Lo. Quantitative analysis of the transrenal excretion of circulating EBV DNA in nasopharyngeal carcinoma patients. *Clinical Cancer Research*, 14(15):4809, 2008.

[42] Karin Birkenkamp-Demtröder, Iver Nordentoft, Emil Christensen, Søren Høyer, Thomas Reinert, Søren Vang, Michael Borre, Mads Agerbæk, Jørgen Bjerggaard Jensen, Torben F. Ørntoft, and Lars Dyrskjøt. Genomic Alterations in Liquid Biopsies from Patients with Bladder Cancer. *European Urology*, 70(1):75–82, 2016.

[43] Ying-Hsiu Su, Mengjun Wang, Dean E Brenner, Alan Ng, Hovsep Melkonyan, Samuil Umansky, Sapna Syngal, and Timothy M Block. Human Urine Contains Small, 150 to 250 Nucleotide- Sized, Soluble DNA Derived from the Circulation and May Be Useful in the Detection of Colorectal Cancer. *Journal of Molecular Diagnostics*, 6(2):101–7, 2004.

[44] Yuxuan Wang, Simeon Springer, Ming Zhang, K Wyatt McMahon, Isaac Kinde, Lisa Dobbyn, Janine Ptak, Henry Brem, Kaisorn Chaichana, Gary L Gallia, Ziya L Gokaslan, Mari L Groves, George I Jallo, Michael Lim, Alessandro Olivi, Alfredo Quinones-Hinojosa, Daniele Rigamonti, Greg J Riggins, Daniel M Sciubba, Jon D Weingart, Jean-Paul Wolinsky, Xiaobu Ye, Sueli Mieko Oba-Shinjo, Suely K N Marie, Matthias Holdhoff, Nishant Agrawal, Luis A Diaz, Nickolas

Papadopoulos, Kenneth W Kinzler, Bert Vogelstein, and Chetan Bettegowda. Detection of tumor-derived DNA in cerebrospinal fluid of patients with primary tumors of the brain and spinal cord. *Proceedings of the National Academy of Sciences*, 112(31):9704–9, 7 2015.

[45] W. Pan, W. Gu, S. Nagpal, M. H. Gephart, and S. R. Quake. Brain Tumor Mutations Detected in Cerebral Spinal Fluid. *Clinical Chemistry*, 61(3):514–522, 2015.

[46] Leticia De Mattos-Arruda, Regina Mayor, Charlotte K. Y. Ng, Britta Weigelt, Francisco Martínez-Ricarte, Davis Torrejon, Mafalda Oliveira, Alexandra Arias, Carolina Raventos, Jiabin Tang, Elena Guerini-Rocco, Elena Martínez-Sáez, Sergio Lois, Oscar Marín, Xavier de la Cruz, Salvatore Piscuoglio, Russel Towers, Ana Vivancos, Vicente Peg, Santiago Ramon y Cajal, Joan Carles, Jordi Rodon, María González-Cao, Josep Tabernero, Enriqueta Felip, Joan Sahuquillo, Michael F. Berger, Javier Cortes, Jorge S. Reis-Filho, and Joan Seoane. Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. *Nature Communications*, 6:8839, 2015.

[47] Krishna B Sriram, Vandana Relan, Belinda E Clarke, Edwina E Duhig, Morgan N Windsor, Kevin S Matar, Rishendran Naidoo, Linda Passmore, Elizabeth McCaul, Deborah Courtney, Ian A Yang, Rayleen V Bowman, and Kwun M Fong. Pleural fluid cell-free DNA integrity index to identify cytologically negative malignant pleural effusions including mesotheliomas. *BMC Cancer*, 12(1):428, 2012.

[48] Suhail K. Mithani, Ian M. Smith, Shaoyu Zhou, Andrew Gray, Wayne M. Koch, Anirban Maitra, and Joseph A. Califano. Mitochondrial resequencing arrays detect tumor-specific mutations in salivary rinses of patients with head and neck cancer. *Clinical Cancer Research*, 13(24):7335–7340, 2007.

[49] Matthew W. Snyder, Martin Kircher, Andrew J. Hill, Riza M. Daza, and Jay Shendure. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell*, 164:57–68, 2016.

[50] Roni Lehmann-Werman, Daniel Neiman, Hai Zemmour, Joshua Moss, Judith Magenheim, Adi Vaknin-Dembinsky, Sten Rubertsson, Bengt Nellgård, Kaj Blennow, Henrik Zetterberg, Kirsty Spalding, Michael J. Haller, Clive H. Wasserfall, Desmond A. Schatz, Carla J. Greenbaum, Craig Dorrell, Markus Grompe, Aviad Zick, Ayala Hubert, Myriam Maoz, Volker Fendrich, Detlef K Bartsch, Talia Golan, Shmuel A. Ben Sasson, Gideon Zamir, Aharon Razin, Howard Cedar, A. M. James Shapiro, Benjamin Glaser, Ruth Shemer, and Yuval Dor. Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proceedings of the National Academy of Sciences*, 113(13):1826–34, 2016.

[51] Kun Sun, Peiyong Jiang, K. C. Allen Chan, John Wong, Yvonne K. Y. Cheng, Raymond H. S. Liang, Wai-kong Chan, Edmond S. K. Ma, Stephen L. Chan, Suk Hang Cheng, Rebecca W. Y. Chan, Yu K. Tong, Simon S. M. Ng, Raymond

S. M. Wong, David S. C. Hui, Tse Ngong Leung, Tak Y. Leung, Paul B. S. Lai, Rossa W. K. Chiu, and Yuk Ming Dennis Lo. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. *Proceedings of the National Academy of Sciences*, 112(40):E5503–E5512, 2015.

[52] Yanni Y N Lui, Ki Wai Chik, Rossa W K Chiu, Cheong Yip Ho, Christopher W K Lam, and Y. M Dennis Lo. Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. *Clinical Chemistry*, 48(3):421–427, 2002.

[53] Suzan Tug, Susanne Helmig, Eva Ricarda Deichmann, Anna Schmeier-Jürchott, Eva Wagner, Tim Zimmermann, Markus Radsak, Mauro Giacca, and Perikles Simon. Exercise-induced increases in cell free DNA in human plasma originate predominantly from cells of the haematopoietic lineage. *Exercise Immunology Review*, 21(27):164–173, 2015.

[54] Edward W H To, K C Allen Chan, Sing-fai Leung, Sing-fai Leung, Lisa Y S Chan, Ka-fai To, Anthony T C Chan, Philip J Johnson, and Y M Dennis Lo. Rapid Clearance of Plasma Epstein-Barr Virus DNA After Surgical Treatment of Nasopharyngeal Carcinoma. *Clinical Cancer Research*, 9:3254–3259, 2003.

[55] Y M Lo, J Zhang, T N Leung, T K Lau, A M Chang, and N M Hjelm. Rapid clearance of fetal DNA from maternal plasma. *American Journal of Human Genetics*, 64(1):218–24, 1999.

[56] Svetlana N. Tamkovich, Anna V. Cherepanova, Elena V. Kolesnikova, Elena Y. Rykova, Dmitrii V. Pyshnyi, Valentin V. Vlassov, and Pavel P. Laktionov. Circulating DNA and DNase Activity in Human Blood. *Annals of the New York Academy of Sciences*, 1075:191–196, 9 1.

[57] Nancy B Y Tsui, Peiyong Jiang, Katherine C K Chow, Xiaoxi Su, Tak Y. Leung, Hao Sun, K. C Allen Chan, Rossa W K Chiu, and Y. M Dennis Lo. High Resolution Size Analysis of Fetal DNA in the Urine of Pregnant Women by Paired-End Massively Parallel Sequencing. *PLoS ONE*, 7(10):e48319, 2012.

[58] T M Chused, A D Steinberg, and N Talal. The clearance and localization of nucleic acids by New Zealand and normal mice. *Clinical and Experimental Immunology*, 12(4):465–76, 1972.

[59] Jonathan C.M. Wan, Charles Massie, Javier Garcia-Corbacho, Florent Mouliere, James D Brenton, Carlos Caldas, Simon Pacey, Richard Baird, and Nitzan Rosenfeld. Liquid biopsies come of age: Towards implementation of circulating tumour DNA, 2 2017.

[60] M B Giacona, G C Ruben, K A Iczkowski, T B Roos, D M Porter, and G D Sorenson. Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls. *Pancreas*, 17(1):89–97, 7 1998.

[61] Y M Dennis Lo, K C Allen Chan, Hao Sun, Eric Z Chen, Peiyong Jiang, Fiona M F Lun, Yama W Zheng, Tak Y Leung, Tze K Lau, Charles R Cantor, and Rossa W K Chiu. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Science Transl. Med.*, 2(61):61ra91, 2010.

[62] Alain R. Thierry, Florent Mouliere, Celine Gongora, Jeremy Ollier, Bruno Robert, Marc Ychou, Maguy del Rio, and Franck Molina. Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. *Nucleic Acids Research*, 38(18):6159–6175, 2010.

[63] D. Nadano, T. Yasuda, and K. Kishi. Measurement of deoxyribonuclease I activity in human tissues and body fluids by a single radial enzyme-diffusion method. *Clinical Chemistry*, 39(3):448–452, 1993.

[64] Florent Mouliere, Safia El Messaoudi, Celine Gongora, Anne-Sophie Guedj, Bruno Robert, Maguy Del Rio, Franck Molina, Pierre-Jean Lamy, Evelyne Lopez-Crapez, Muriel Mathonnet, Marc Ychou, Denis Pezet, and Alain R Thierry. Circulating Cell-Free DNA from Colorectal Cancer Patients May Reveal High KRAS or BRAF Mutation Load. *Translational Oncology*, 6(3):319–328, 2013.

[65] Hunter R. HR Underhill, Jacob O. Kitzman, Sabine Hellwig, Noah C. Welker, Riza RM Daza, Daniel N. Baker, Keith M. Gligorich, Robert C. RC Rostomily, Mary P. Bronner, and Jay Shendure. Fragment Length of Circulating Tumor DNA. *PLOS Genet*, 12(7):426–37, 2016.

[66] Tobias M Gorges, Johanna Schiller, Arndt Schmitz, Daniel Schuetzmann, Christoph Schatz, Thomas M Zollner, Thomas Krahn, and Oliver von Ahsen. Cancer therapy monitoring in xenografts by quantitative analysis of circulating tumor DNA. *Biomarkers*, 17(6):498–506, 2012.

[67] Yama W L Zheng, K. C Allen Chan, Hao Sun, Peiyong Jiang, Xiaoxi Su, Eric Z. Chen, Fiona M F Lun, Emily C W Hung, Vincent Lee, John Wong, Paul B S Lai, Chi Kong Li, Rossa W K Chiu, and Y. M Dennis Lo. Nonhematopoietically derived DNA is shorter than hematopoietically derived DNA in plasma: A transplantation model. *Clinical Chemistry*, 58(3):549–558, 2012.

[68] Peiyong Jiang and Y. M Dennis Lo. The Long and Short of Circulating Cell-Free DNA and the Ins and Outs of Molecular Diagnostics. *Trends in Genetics*, 32(6):360–371, 2016.

[69] Suk Hang Cheng, Peiyong Jiang, Kun Sun, Yvonne K Y Cheng, K. C Allen Chan, Tak Y. Leung, Rossa W K Chiu, and Y. M Dennis Lo. Noninvasive prenatal testing by nanopore sequencing of maternal plasma DNA: Feasibility assessment. *Clinical Chemistry*, 61(10):1305–1306, 2015.

[70] Martin Beránek, Igor Sirák, Milan Vošmik, Jiří Petera, Monika Drastíková, and Vladimír Palička. Carrier molecules and extraction of circulating tumor DNA for next generation sequencing in colorectal cancer. *Acta Medica*, 59(2):54–58, 2016.

[71] Marie-Theres Gansauge and Matthias Meyer. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols*, 8(4):737–48, 2013.

[72] Philip Burnham, Min Seong Kim, Sean Agbor-Enoh, Helen Luikart, Hannah A Valantine, Kiran K Khush, and Iwijn De Vlaminck. Single-stranded DNA library preparation uncovers the origin and diversity of ultrashort cell-free DNA in plasma. *Scientific Reports*, 6(February):27859, 6 2016.

[73] S. V. Kostyuk, E. M. Malinovskaya, a. V. Ermakov, T. D. Smirnova, L. V. Kameneva, O. V. Chvartatskaya, P. a. Loseva, E. S. Ershova, L. N. Lyubchenko, and N. N. Veiko. Fragments of cell-free DNA increase transcription in human mesenchymal stem cells, activate TLR-dependent signal pathway, and suppress apoptosis. *Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry*, 6(1):68–74, 2012.

[74] Sachiko Nishimoto, Daiju Fukuda, Yasutomi Higashikuni, Kimie Tanaka, Yoichiro Hirata, Chie Murata, Fukiko Sato, Masahiro Bando, Shusuke Yagi, Takeshi Soeki, Tetsuya Hayashi, Issei Imoto, Joo-Ri Kim-Kaneyama, Fukiko Sato, Masahiro Bando, Shusuke Yagi, Takeshi Soeki, Tetsuya Hayashi, Issei Imoto, Hiroshi Sakaue, Michio Shimabukuro, and Masataka Sata. Obesity-induced DNA released from adipocytes stimulates chronic adipose tissue inflammation and insulin resistance. *Science Advances*, 2(March):1–12, 2016.

[75] Jerker Porath and Stanley M Gartler. Cellular Uptake of Deoxyribonucleic Acid by Human Tissue Culture Cells. *Nature*, 184(4697):1505–1506, 11 1959.

[76] M Dvořáková, V Karafiát, P Pajer, E Kluzáková, K Jarkovská, S Peková, L Krutílková, M Dvořák, M Dvor˘a´kova´, V Karafiát, P Pajer, M Dvořáková, E Kluzáková, K Jarkovská, S Peková, L Krutílková, and M Dvořák. DNA released by leukemic cells contributes to the disruption of the bone marrow microenvironment. *Oncogene*, 32(44):5201–5209, 2013.

[77] Dolores C. García-Olmo, Carolina Domínguez, Mariano García-Arranz, Phillipe Anker, Maurice Stroun, José M. García-Verdugo, and Damián García-Olmo. Cell-free nucleic acids circulating in the plasma of colorectal cancer patients induce the oncogenic transformation of susceptible cultured cells. *Cancer Research*, 70(2):560–567, 2010.

[78] Young Seok Ju, Jose M C Tubio, William Mifsud, Beiyuan Fu, Helen R. Davies, Manasa Ramakrishna, Yilong Li, Lucy Yates, Gunes Gundem, Patrick S. Tarpey, Sam Behjati, Elli Papaemmanuil, Sancha Martin, Anthony Fullam, Moritz Gerstung, Jyoti Nangalia, Anthony R. Green, Carlos Caldas, Åke Borg, Andrew Tutt, Ming Ta Michael Lee, Laura J. Van'T Veer, Benita K T Tan, Samuel Aparicio, Paul N. Span, John W M Martens, Stian Knappskog, Anne Vincent-Salomon, Anne Lise Børresen-Dale, Jórunn Erla Eyfjörd, Adrienne M. Flanagan, Christopher Foster, David E. Neal, Colin Cooper, Rosalind Eeles, Sunil R. Lakhani, Christine Desmedt, Gilles Thomas, Andrea L. Richardson, Colin A. Purdie, Alastair M. Thompson, Ultan McDermott, Fengtang Yang, Serena Nik-Zainal, Peter J. Campbell, and Michael R. Stratton. Frequent somatic transfer

of mitochondrial DNA into the nuclear genome of human cancer cells. *Genome Research*, 25(6):814–824, 2015.

[79] M Willems, H Moshage, F Nevens, J Fevery, and S H Yap. Plasma collected from heparinized blood is not suitable for HCV-RNA detection by conventional RT-PCR assay. *Journal of Virological Methods*, 42(1):127–130, 4 1993.

[80] Yanan Kuang, Andrew Rogers, Beow Y Yeap, Lilin Wang, Mike Makrigiorgos, Sara Thiede, Robert J Distel, and Pasi A Jaänne. Noninvasive Detection of EGFR T790M in Gefitinib or Erlotinib Resistant Non-Small Cell Lung Cancer. *Clinical Cancer Research*, 15(8):2630–2636, 2009.

[81] Monika Jung, Silke Klotzek, Michaela Lewandowski, Michael Fleischhacker, and Klaus Jung. Changes in Concentration of DNA in Serum and Plasma during Storage of Blood Samples. *Clinical Chemistry*, 49(6):1028–1029, 2003.

[82] Sonya Parpart-Li, Bjarne Bartlett, Maria Popoli, Vilmos Adleff, Laura Tucker, Rebecca Steinberg, Andrew Georgiadis, Jillian Phallen, Julie R Brahmer, Nilofer A Azad, Ilene Browner, Daniel A Laheru, Victor E Velculescu, Mark Sausen, and Luis A Diaz. The effect of preservative and temperature on the analysis of circulating tumor DNA. *Clin. Cancer Res.*, 23(10):2471–2477, 11 2017.

[83] Safia El Messaoudi, Fanny Rolet, Florent Mouliere, and Alain R. Thierry. Circulating cell free DNA: Preanalytical considerations. *Clinica Chimica Acta*, 424:222–230, 2013.

[84] Patricia Valda Toro, Bracha Erlanger, Julia a. Beaver, Rory L. Cochran, Dustin a. VanDenBerg, Elizabeth Yakim, Karen Cravero, David Chu, Daniel J. Zabransky, Hong Yuen Wong, Sarah Croessmann, Heather Parsons, Paula J. Hurley, Josh Lauring, and Ben Ho Park. Comparison of cell stabilizing blood collection tubes for circulating plasma tumor DNA. *Clinical Biochemistry*, 48(15):993–998, 2015.

[85] Qing Kang, N. Lynn Henry, Costanza Paoletti, Hui Jiang, Pankaj Vats, Arul M. Chinnaiyan, Daniel F. Hayes, Sofia D. Merajver, James M. Rae, and Muneesh Tewari. Comparative Analysis of Circulating Tumor DNA Stability In K3EDTA, Streck and CellSave Blood Collection Tubes. *Clinical Biochemistry*, 49(18):1354–1360, 2016.

[86] Inga Medina Diaz, Annette Nocon, Daniel H. Mehnert, Johannes Fredebohm, Frank Diehl, and Frank Holtrup. Performance of Streck cfDNA blood collection tubes for liquid biopsy testing. *PLoS ONE*, 11(11):e0166354, 2016.

[87] S. E. Norton, J. M. Lechner, T. Williams, and M. R. Fernando. A stabilizing reagent prevents cell-free DNA contamination by cellular DNA in plasma during blood sample storage and shipping as determined by digital PCR. *Clinical Biochemistry*, 46(15):1561–1565, 2013.

[88] Steffen Dietz, Uwe Schirmer, Clementine Merce, Nikolas von Bubnoff, Edgar Dahl, Michael Meister, Thomas Muley, Michael Thomas, and Holger Sultmann. Low Input Whole-Exome Sequencing to Determine the Representation of the Tumor Exome in Circulating DNA of Non-Small Cell Lung Cancer Patients. *PLoS ONE*, 11(8):e0161012, 2016.

[89] QIAGEN. therascreen EGFR RGQ PCR Kit Handbook - Version 1, 2013.

[90] U.S. Food and Drug Administration. Premarket Approval P150044 - Cobas EGFR MUTATION TEST V2, 2016.

[91] Tony K F Yung, K. C Allen Chan, Tony S K Mok, Joanna Tong, Ka Fai To, and Y. M Dennis Lo. Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non-small cell lung cancer patients. *Clinical Cancer Research*, 15(6):2076–2084, 2009.

[92] Gaia Schiavon, Sarah Hrebien, I. Garcia-Murillas, Rosalind J Cutts, Alex Pearson, Noelia Tarazona, Kerry Fenwick, Iwanka Kozarewa, E. Lopez-Knowles, Ricardo Ribas, Ashutosh Nerurkar, Peter Osin, Sarat Chandarlapaty, L.-A. Martin, Mitch Dowsett, Ian E Smith, and Nicholas C Turner. Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer. *Science Translational Medicine*, 7(313):182–313, 2015.

[93] Valerie Taly, Deniz Pekin, Leonor Benhaim, Steve K. Kotsopoulos, Delphine Le Corre, Xinyu Li, Ivan Atochin, Darren R. Link, Andrew D. Griffiths, Karine Pallier, Hélène Blons, Olivier Bouché, Bruno Landi, J. Brian Hutchison, and Pierre Laurent-Puig. Multiplex picodroplet digital PCR to detect KRAS mutations in circulating DNA from the plasma of colorectal cancer patients. *Clinical Chemistry*, 59(12):1722–1731, 2013.

[94] Adrian G. Sacher, Cloud Paweletz, Suzanne E. Dahlberg, Ryan S. Alden, Allison O'Connell, Nora Feeney, Stacy L. Mach, Pasi A. Jänne, and Geoffrey R. Oxnard. Prospective Validation of Rapid Plasma Genotyping for the Detection of EGFR and KRAS Mutations in Advanced Lung Cancer. *JAMA Oncology*, 2(8):1014–22, 2016.

[95] Andre Marziali, Joel Pel, Dan Bizzotto, and Lorne A. Whitehead. Novel electrophoresis mechanism based on synchronous alternating drag perturbation. *Electrophoresis*, 26(1):82–90, 2005.

[96] Jason D. Thompson, Gosuke Shibahara, Sweta Rajan, Joel Pel, and Andre Marziali. Winnowing DNA for rare sequences: Highly specific sequence and methylation based enrichment. *PLoS ONE*, 7(2):e31597, 2012.

[97] Jin Li, Lilin Wang, Harvey Mamon, Matthew H Kulke, Ross Berbeco, and G Mike Makrigiorgos. Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing. *Nature Medicine*, 14(5):579–584, 2008.

[98] Minakshi Guha, Elena Castellanos-Rizaldos, Pingfang Liu, Harvey Mamon, and G. Mike Makrigiorgos. Differential strand separation at critical temperature: A minimally disruptive enrichment method for low-abundance unknown DNA mutations. *Nucleic Acids Research*, 41(3):e50, 2013.

[99] M. W. Schmitt, Scott R. Kennedy, Jesse J. Salk, Edward J. Fox, Joseph B. Hiatt, and Lawrence A. Loeb. Detection of ultra-rare mutations by next-generation sequencing. *Proceedings of the National Academy of Sciences*, 109(36):14508–14513, 2007.

[100] K. C. A. Chan, P. Jiang, Y. W. L. Zheng, G. J. W. Liao, H. Sun, J. Wong, S. S. N. Siu, W. C. Chan, S. L. Chan, a. T. C. Chan, P. B. S. Lai, R. W. K. Chiu, and Y. M. D. Lo. Cancer Genome Scanning in Plasma: Detection of Tumor-Associated Copy Number Aberrations, Single-Nucleotide Variants, and Tumoral Heterogeneity by Massively Parallel Sequencing. *Clinical Chemistry*, 59(1):211–224, 2013.

[101] Alain R Thierry, Florent Mouliere, Safia El Messaoudi, Caroline Mollevi, Evelyne Lopez-Crapez, Fanny Rolet, Brigitte Gillet, Celine Gongora, Pierre Dechelotte, Bruno Robert, Maguy Del Rio, Pierre-Jean Lamy, Frederic Bibeau, Michelle Nouaille, Virginie Loriot, Anne-Sophie Jarrousse, Franck Molina, Muriel Mathonnet, Denis Pezet, and Marc Ychou. Clinical validation of the detection of KRAS and BRAF mutations from circulating tumor DNA. *Nature Medicine*, 20(4):430–435, 2014.

[102] Eleonora Zonta, Fanny Garlan, Karla Perez-Toralla, and Valérie Taly. Multiplex Detection of Rare Mutations by Picoliter Droplet Based Digital PCR: Sensitivity and Specificity Considerations. *PLoS ONE*, 11(7):e0159094, 2016.

[103] Evelyn Kidess, Kyra Heirich, Matthew Wiggin, Valentina Vysotskaia, Brendan C Visser, Andre Marziali, Bertram Wiedenmann, Jeffrey A Norton, Mark Lee, Stefanie S Jeffrey, and George A Poultsides. Mutation profiling of tumor DNA from plasma and tumor tissue of colorectal cancer patients with a novel, high-sensitivity multiplexed mutation detection platform. *Oncotarget*, 6(4):2549–61, 2 2015.

[104] Davina Gale, Vincent Plagnol, Andrew Lawson, Michelle Pugh, Sarah Smalley, Karen Howarth, Mikidache Madi, Bradley Durham, Vasudev Kumanduri, Kitty Lo, James Clark, Emma Green, Nitzan Rosenfeld, and Tim Forshew. Abstract 3639: Analytical performance and validation of an enhanced TAm-Seq circulating tumor DNA sequencing assay. *Cancer Research*, 76(14 Supplement):3639, 7 2016.

[105] Richard B. Lanman, Stefanie A. Mortimer, Oliver A. Zill, Dragan Sebisanovic, Rene Lopez, Sibel Blau, Eric A. Collisson, Stephen G. Divers, Dave S B Hoon, E. Scott Kopetz, Jeeyun Lee, Petros G. Nikolinakos, Arthur M. Baca, Bahram G. Kermani, Helmy Eltoukhy, and Amir Ali Talasaz. Analytical and clinical validation of a digital sequencing panel for quantitative, highly accurate evaluation of cell-free circulating tumor DNA. *PLoS ONE*, 10(10):e0140712, 2015.

[106] Aaron M. Newman, Alexander F. Lovejoy, Daniel M. Klass, David M. Kurtz, Jacob J. Chabon, Florian Scherer, H Stehr, C L Liu, S V Bratman, C Say, L Zhou, J N Carter, R B West, G W Sledge Jr., J B Shrager, B W Loo Jr., J W Neal, H A Wakelee, M Diehn, and A A Alizadeh. Integrated digital error suppression for improved detection of circulating tumor DNA. *Nature Biotechnology*, 34(5):547–55, 2016.

[107] Maria Schwaederle, Hatim Husain, Paul T. Fanta, David E. Piccioni, Santosh Kesari, Richard B. Schwab, Sandip P. Patel, Olivier Harismendy, Megumi Ikeda, Barbara A. Parker, and Razelle Kurzrock. Use of Liquid Biopsies in Clinical Oncology: Pilot Experience in 168 Patients. *Clinical Cancer Research*, 22(22):5497–5505, 2016.

[108] Jean-Sebastien Frenel, Suzanne Carreira, Jane Goodall, Desamparados Roda Perez, Raquel Perez Lopez, Nina Tunariu, Ruth Riisnaes, Susana Miranda, Ines Figueiredo, Daniel Nava Rodrigues, Alan Smith, Christophe Leux, Isaac Garcia-Murillas, Roberta Ferraldeschi, David Lorente, Joaquin Mateo, Michael Ong, Timothy a. Yap, Udai Banerji, Delila Gasi Tandefelt, Nick Turner, Gerhardt Attard, and Johann S. de Bono. Serial Next Generation Sequencing of Circulating Cell Free DNA Evaluating Tumour Clone Response To Molecularly Targeted Drug Administration. *Clinical Cancer Research*, 21(20):4586–96, 2015.

[109] Ronald Lebofsky, Charles Decraene, Virginie Bernard, Maud Kamal, Anthony Blin, Quentin Leroy, Thomas Rio Frio, Gaëlle Pierron, Céline Callens, Ivan Bieche, Adrien Saliou, Jordan Madic, Etienne Rouleau, François-Clément Bidard, Olivier Lantz, Marc-Henri Stern, Christophe Le Tourneau, and Jean-Yves Pierga. Circulating tumor DNA as a non-invasive substitute to metastasis biopsy for tumor genotyping and personalized medicine in a prospective trial across all tumor types. *Molecular Oncology*, 9(4):783–790, 2015.

[110] Ellen Heitzer, Peter Ulz, Jelena Belic, Stefan Gutschi, Franz Quehenberger, Katja Fischereder, Theresa Benezeder, Martina Auer, Carina Pischler, Sebastian Mannweiler, Martin Pichler, Florian Eisner, Martin Haeusler, Sabine Riethdorf, Klaus Pantel, Hellmut Samonigg, Gerald Hoefler, Herbert Augustin, Jochen B Geigl, and Michael R Speicher. Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. *Genome Medicine*, 5(4):30, 2013.

[111] Aaron M Newman, Scott V Bratman, Jacqueline To, Jacob F Wynne, Neville C W Eclov, Leslie a Modlin, Chih Long Liu, Joel W Neal, Heather a Wakelee, Robert E Merritt, Joseph B Shrager, Billy W Loo, Ash a Alizadeh, and Maximilian Diehn. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nature Medicine*, 20(5):548–54, 2014.

[112] Isaac Kinde, Jian Wu, Nick Papadopoulos, Kenneth W Kinzler, and Bert Vogelstein. Detection and quantification of rare mutations with massively parallel sequencing. *Proceedings of the National Academy of Sciences*, 108(23):9530–9535, 2011.

[113] Sarah B. Ng, Clarinda Chua, Matthew Ng, Anna Gan, Polly SY Poon, Melissa Teo, Cherylin Fu, Wei Qiang Leow, Kiat Hon Lim, Alexander Chung, Si-Lin Koo, Su Pin Choo, Danliang Ho, Steve Rozen, Patrick Tan, Mark Wong, William F. Burkholder, and Iain Beehuat Tan. Individualised multiplexed circulating tumour DNA assays for monitoring of tumour presence in patients after colorectal cancer surgery. *Scientific Reports*, 7:40737, 2017.

[114] J. Belic, M. Koch, P. Ulz, M. Auer, T. Gerhalter, S. Mohan, K. Fischereder, E. Petru, T. Bauernhofer, J. B. Geigl, M. R. Speicher, and E. Heitzer. Rapid Identification of Plasma DNA Samples with Increased ctDNA Levels by a Modified FAST-SeqS Approach. *Clinical Chemistry*, 61(6):838–849, 2015.

[115] Viktor A Adalsteinsson, Gavin Ha, Samuel S Freeman, Atish D Choudhury, Daniel G Stover, Heather A Parsons, Gregory Gydush, Sarah C Reed, Denisse Rotem, Justin Rhoades, Denis Loginov, Dimitri Livitz, Daniel Rosebrock, Ignaty Leshchiner, Jaegil Kim, Chip Stewart, Mara Rosenberg, Joshua M Francis, Cheng-Zhong Zhang, Ofir Cohen, Coyin Oh, Huiming Ding, Paz Polak, Max Lloyd, Sairah Mahmud, Karla Helvie, Margaret S Merrill, Rebecca A Santiago, Edward P O'Connor, Seong H Jeong, Rachel Leeson, Rachel M Barry, Joseph F Kramkowski, Zhenwei Zhang, Laura Polacek, Jens G Lohr, Molly Schleicher, Emily Lipscomb, Andrea Saltzman, Nelly M Oliver, Lori Marini, Adrienne G Waks, Lauren C Harshman, Sara M Tolaney, Eliezer M Van Allen, Eric P Winer, Nancy U Lin, Mari Nakabayashi, Mary-Ellen Taplin, Cory M Johannessen, Levi A Garraway, Todd R Golub, Jesse S Boehm, Nikhil Wagle, Gad Getz, J Christopher Love, and Matthew Meyerson. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nature Communications*, 8(1):1324, 2017.

[116] Eser Kirkizlar, Bernhard Zimmermann, Tudor Constantin, Ryan Swenerton, Bin Hoang, Nicholas Wayham, Joshua E Babiarz, Zachary Demko, Robert J Pelham, Stephanie Kareht, Alexander L Simon, Kristine N Jinnett, Matthew Rabinowitz, Styrmir Sigurjonsson, and Matthew Hill. Detection of Clonal and Subclonal Copy-Number Variants in Cell-Free DNA from Patients with Breast Cancer Using a Massively Multiplexed PCR Methodology. *Translational Oncology*, 8(5):407–416, 2015.

[117] Christopher Abbosh, Nicolai J. Birkbak, Gareth A. Wilson, Mariam Jamal-Hanjani, Tudor Constantin, Raheleh Salari, John Le Quesne, David A Moore, Selvaraju Veeriah, Rachel Rosenthal, Teresa Marafioti, Eser Kirkizlar, Thomas B.K. Watkins, Nicholas McGranahan, Sophia Ward, Luke Martinson, Joan Riley, Francesco Fraioli, Maise Al Bakir, Eva Grönroos, Francisco Zambrana, Raymondo Endozo, Wenya Linda Bi, Fiona M. Fennessy, Nicole Sponer, Diana Johnson, Joanne Laycock, Seema Shafi, Justyna Czyzewska-Khan, Andrew Rowan, Tim Chambers, Nik Matthews, Samra Turajlic, Crispin Hiley, Siow Ming Lee, Martin D. Forster, Tanya Ahmad, Mary Falzon, Elaine Borg, David Lawrence, Martin Hayward, Shyam Kolvekar, Nikolaos Panagiotopoulos, Sam M. Janes, Ricky Thakrar, Asia Ahmed, Fiona Blackhall, Yvonne Summers, Dina Hafez, Ashwini Naik, Apratim Ganguly, Stephanie Kareht, Rajesh Shah, Leena Joseph, Anne Marie Quinn, Phil A. Crosbie, Babu Naidu, Gary Middleton,

Gerald Langman, Simon Trotter, Marianne Nicolson, Hardy Remmen, Keith Kerr, Mahendran Chetty, Lesley Gomersall, Dean A. Fennell, Apostolos Nakas, Sridhar Rathinam, Girija Anand, Sajid Khan, Peter Russell, Veni Ezhil, Babikir Ismail, Melanie Irvin-Sellers, Vineet Prakash, Jason F. Lester, Malgorzata Kornaszewska, Richard Attanoos, Haydn Adams, Helen Davies, Dahmane Oukrif, Ayse U. Akarca, John A. Hartley, Helen L. Lowe, Sara Lock, Natasha Iles, Harriet Bell, Yenting Ngai, Greg Elgar, Zoltan Szallasi, Roland F. Schwarz, Javier Herrero, Aengus Stewart, Sergio A. Quezada, Karl S. Peggs, Peter Van Loo, Caroline Dive, C. Jimmy Lin, Matthew Rabinowitz, Hugo J.W.L. Aerts, Allan Hackshaw, Jacqui A. Shaw, Bernhard G. Zimmermann, and Charles Swanton. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature*, 545(7655):446–451, 2017.

[118] Christine A. Parkinson, Davina Gale, Anna M. Piskorz, Heather Biggs, Charlotte Hodgkin, Helen Addley, Sue Freeman, Penelope Moyle, Evis Sala, Karen Sayal, Karen Hosking, Ioannis Gounaris, Mercedes Jimenez-Linan, Helena M. Earl, Wendi Qian, Nitzan Rosenfeld, and James D. Brenton. Exploratory analysis of TP53 mutations in circulating tumour DNA as biomarkers of treatment response for patients with relapsed high-grade serous ovarian carcinoma: a retrospective study. *PLoS Medicine*, 13(12):e1002198, 2016.

[119] Michael J. Overman, Janhavi Modak, Scott Kopetz, Ravi Murthy, James C. Yao, Marshall E. Hicks, James L. Abbruzzese, and Alda L. Tam. Use of research biopsies in clinical trials: Are risks and benefits adequately discussed? *Journal of Clinical Oncology*, 31(1):17–22, 2013.

[120] Paul A. VanderLaan, Norihiro Yamaguchi, Erik Folch, David H. Boucher, Michael S. Kent, Sidharta P. Gangadharan, Adnan Majid, Michael A. Goldstein, Mark S. Huberman, Olivier N. Kocher, and Daniel B. Costa. Success and failure rates of tumor genotyping techniques in routine pathological samples with non-small-cell lung cancer. *Lung Cancer*, 84(1):39–44, 2014.

[121] Peter M. Ellis, Frances A. Shepherd, Michael Millward, Francesco Perrone, Lesley Seymour, Geoffrey Liu, Sophie Sun, Byoung Chul Cho, Alessandro Morabito, Natasha B. Leighl, Martin R. Stockler, Christopher W. Lee, Rafal Wierzbicki, Victor Cohen, Normand Blais, Randeep S. Sangha, Adolfo G. Favaretto, Jin Hyoung Kang, Ming Sound Tsao, Carolyn F. Wilson, Zelanna Goldberg, Keyue Ding, Glenwood D. Goss, and Penelope Ann Bradbury. Dacomitinib compared with placebo in pretreated patients with advanced or metastatic non-small-cell lung cancer (NCIC CTG BR.26): A double-blind, randomised, phase 3 trial. *Lancet Oncol.*, 15(12):1379–1388, 2014.

[122] Helmut H. Popper. Commentary on tumor heterogeneity. *Translational Lung Cancer Research*, 5(4):433–435, 2016.

[123] L. De Mattos-Arruda, B. Weigelt, J. Cortes, H. H. Won, C. K. Y. Ng, P. Nuciforo, F-C F.-C. Bidard, C. Aura, C. Saura, V. Peg, S. Piscuoglio, M. Oliveira, Y. Smolders, P. Patel, L. Norton, J. Tabernero, M. F. Berger, J. Seoane, and J. S. Reis-Filho. Capturing Intra-Tumor Genetic Heterogeneity by De Novo

Mutation Profiling of Circulating Cell-Free Tumor DNA: A Proof-of-Principle. *Annals of Oncology*, 25(July):1729–1735, 2014.

[124] M. Jamal-Hanjani, G. A. Wilson, S. Horswell, R. Mitter, O. Sakarya, T. Constantin, R. Salari, E. Kirkizlar, S. Sigurjonsson, R. Pelham, S. Kareht, B. Zimmermann, and C. Swanton. Detection of Ubiquitous and Heterogeneous Mutations in Cell-Free DNA from Patients with Early-Stage Non–Small-Cell Lung Cancer. *Ann. Oncol.*, 27(5):862–7, 2016.

[125] Muhammed Murtaza, Sarah-jane Dawson, Katherine Pogrebniak, Oscar M Rueda, Elena Provenzano, John Grant, Suet-feung Chin, Dana W Y Tsui, Francesco Marass, Davina Gale, H Raza Ali, Pankti Shah, Tania Contente-cuomo, Hossein Farahani, Karey Shumansky, Zoya Kingsbury, Sean Humphray, David Bentley, Sohrab P Shah, Matthew Wallis, Nitzan Rosenfeld, and Carlos Caldas. Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nature Communications*, 6:8760, 2015.

[126] Michael J. Duffy. Serum tumor markers in breast cancer: Are they of clinical value? *Clinical Chemistry*, 52(3):345–351, 2006.

[127] Reza Fazel, Harlan M Krumholz, Yongfei Wang, Joseph S Ross, Jersey Chen, Henry H Ting, Nilay D Shah, Khurram Nasir, Andrew J Einstein, and Brahmajee K Nallamothu. Exposure to Low-Dose Ionizing Radiation from Medical Imaging Procedures. *New England Journal of Medicine*, 361(9):849–857, 8 2009.

[128] Aparna A. Kamat, Farideh Z. Bischoff, Dianne Dang, Matthew F. Baldwin, Liz Y. Han, Yvonne G. Lin, William M. Merritt, Charles N. Landen, Chunhua Lu, David M. Gershenson, Joe L. Simpson, and Anil K. Sood. Circulating cell-free DNA: A novel biomarker for response to therapy in ovarian carcinoma. *Cancer Biology and Therapy*, 5(10):1369–1374, 2006.

[129] Chetan Bettegowda, Mark Sausen, Rebecca J. Leary, Isaac Kinde, Yuxuan Wang, Nishant Agrawal, Bjarne R. Bartlett, Hao Wang, Brandon Luber, Rhoda M. Alani, Emmanuel S. Antonarakis, Nilofer S. Azad, Alberto Bardelli, Henry Brem, John L. Cameron, Clarence C. Lee, Leslie a. Fecher, Gary L. Gallia, Peter Gibbs, Dung Le, Robert L. Giuntoli, Michael Goggins, Michael D. Hogarty, Matthias Holdhoff, Seung-Mo S.-M. Hong, Yuchen Jiao, Hartmut H. Juhl, Jenny J. Kim, Giulia Siravegna, Daniel a. Laheru, Calogero Lauricella, Michael Lim, Evan J. Lipson, Suely Kazue Nagahashi Marie, George J. Netto, Kelly S. Oliner, Alessandro Olivi, Louise Olsson, Gregory J. Riggins, Andrea Sartore-Bianchi, Kerstin Schmidt, Le-Ming Shih, Sueli Mieko Oba-Shinjo, Salvatore Siena, Dan Theodorescu, Jeanne Tie, Timothy T. Harkins, Silvio Veronese, T.-L. Tian-Li Wang, Jon D. Weingart, Christopher L. Wolfgang, Laura D. Wood, Dongmei Xing, Ralph H. Hruban, Jian Wu, Peter J. Allen, C. Max Schmidt, Michael a. Choti, Victor E. Velculescu, Kenneth W. Kinzler, Bert Vogelstein, Nickolas Papadopoulos, and Luis A. Diaz. Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. *Science Transl. Med.*, 6(224):224ra24, 2 2014.

[130] Anneli Dowler Nygaard, Karen Lise Garm Spindler, Niels Pallisgaard, Rikke Fredslund Andersen, and Anders Jakobsen. The prognostic value of KRAS mutated plasma DNA in advanced non-small cell lung cancer. *Lung Cancer*, 79(3):312–317, 2013.

[131] Shuhang Wang, Tongtong An, Jie Wang, Jun Zhao, Zhijie Wang, Minglei Zhuo, Hua Bai, Lu Yang, Yan Zhang, Xin Wang, Jianchun Duan, Yuyan Wang, Qingzhi Guo, and Meina Wu. Potential clinical significance of a plasma-based KRAS mutation analysis in patients with advanced non-small cell lung cancer. *Clinical Cancer Research*, 16(4):1324–1330, 2010.

[132] O. Gautschi, B. Huegli, A. Ziegler, M. Gugger, J. Heighway, D. Ratschiller, P. C. Mack, P. H. Gumerlock, H. J. Kung, R. a. Stahel, D. R. Gandara, and D. C. Betticher. Origin and prognostic value of circulating KRAS mutations in lung cancer patients. *Cancer Letters*, 254(2):265–273, 2007.

[133] Elin S Gray, Helen Rizos, Anna L Reid, Suzanah C Boyd, Michelle R Pereira, Johnny Lo, Varsha Tembe, James Freeman, Jenny H J Lee, Richard A Scolyer, Kelvin Siew, Chris Lomma, Adam Cooper, Muhammad A Khattak, Tarek M Meniawy, Georgina V Long, Matteo S Carlino, Michael Millward, and Melanie Ziman. Circulating tumor DNA to monitor treatment response and detect acquired resistance in patients with metastatic melanoma. *Oncotarget*, 6(39):42008–18, 2015.

[134] Evan J Lipson, Victor E Velculescu, Theresa S Pritchard, Mark Sausen, Drew M Pardoll, Suzanne L Topalian, and Luis a Diaz. Circulating tumor DNA analysis as a real-time method for monitoring tumor burden in melanoma patients undergoing treatment with immune checkpoint blockade. *Journal for Immunotherapy of Cancer*, 2(1):42, 2014.

[135] Chiya Oshiro, Naofumi Kagara, Yasuto Naoi, Masafumi Shimoda, Atsushi Shimomura, Naomi Maruyama, Kenzo Shimazu, Seung Jin Kim, and Shinzaburo Noguchi. PIK3CA mutations in serum DNA are predictive of recurrence in primary breast cancer patients. *Breast Cancer Research and Treatment*, 150(2):299–307, 2015.

[136] Florian Scherer, David M. Kurtz, Aaron M. Newman, Henning Stehr, Alexander F.M. Craig, Mohammad Shahrokh Esfahani, Alexander F. Lovejoy, Jacob J. Chabon, Daniel M. Klass, Chih Long Liu, Li Zhou, Cynthia Glover, Brendan C. Visser, George A. Poultsides, Ranjana H. Advani, Lauren S. Maeda, Neel K. Gupta, Ronald Levy, Robert S. Ohgami, Christian A. Kunder, Maximilian Diehn, and Ash A. Alizadeh. Distinct biological subtypes and patterns of genome evolution in lymphoma revealed by circulating tumor DNA. *Science Translational Medicine*, 8(364):364ra155, 2016.

[137] European Medicines Agency. Iressa: Public Assessment Report - Product Information, 2016.

[138] U.S. Food and Drug Administration. Summary of Safety and Effectiveness Data (SSED). Technical report, U.S. Food and Drug Administration, 2011.

[139] B. Han, S. Tjulandin, K. Hagiwara, N. Normanno, L. Wulandari, Konstantin Konstantinovich, A. Hudoyo, M. Ratcliffe, R. McCormack, and M. Reck. Determining the prevalence of EGFR mutations in Asian and Russian patients (pts) with advanced non-small-cell lung cancer (aNSCLC) of adenocarcinoma (ADC) and non-ADC histology: IGNITE study. *Annals of Oncology*, 26(Supplement 1):i29–i44, 2015.

[140] J Remon, C Caramella, C Jovelet, L Lacroix, A Lawson, S Smalley, K Howarth, D Gale, N Rosenfeld, E Green, V Plagnol, D Planchard, M V Bluthgen, A Gazzah, C Pannet, C Nicotra, J C Soria, and B Besse. Osimertinib benefit in ctDNA T790M positive, EGFR-mutant NSCLC patients. In *IASLC 17th World Conference on Lung Cancer*, Vienna, Austria, 2016.

[141] Institut Gustave Roussy. The Prospective MOSCATO 01 Trial Demonstrates that Molecular "Portraits" Improve Outcome of Patients with Metastatic Cancer, 2016.

[142] Geoffrey R. Oxnard, Kenneth S. Thress, Ryan S. Alden, Rachael Lawrance, Cloud P. Paweletz, Mireille Cantarini, James Chih-Hsin Yang, J. Carl Barrett, and Pasi A Janne. Association between plasma genotyping and outcomes of treatment with osimertinib (AZD9291) in advanced non-small-cell lung cancer. *Journal of Clinical Oncology*, 34(28):3375–3382, 10 2016.

[143] Michael J. Pishvaian, R. Joseph Bender, Lynn M. Matrisian, Lola Rahib, Andrew Hendifar, William A. Hoos, Sam Mikhail, Vincent Chung, Vincent Picozzi, Craig Heartwell, Kimberly Mason, Katelyn Varieur, Metasebia Aberra, Subha Madhavan, Emanuel Petricoin III, Jonathan R. Brody, Michael J. Pishvaian, R. Joseph Bender, Lynn M. Matrisian, Lola Rahib, Andrew Hendifar, William A. Hoos, Sam Mikhail, Vincent Chung, Vincent Picozzi, Craig Heartwell, Kimberly Mason, Katelyn Varieur, Metasebia Aberra, Subha Madhavan, Emanuel Petricoin III, and Jonathan R. Brody. A pilot study evaluating concordance between blood-based and patient-matched tumor molecular testing within pancreatic cancer patients participating in the Know Your Tumor (KYT) initiative. *Oncotarget*, 5(0), 2016.

[144] Marco Gerlinger, Andrew J. Rowan, Stuart Horswell, James Larkin, David Endesfelder, Eva Gronroos, Pierre Martinez, Nicholas Matthews, Aengus Stewart, Patrick Tarpey, Ignacio Varela, Benjamin Phillimore, Sharmin Begum, Neil Q. McDonald, Adam Butler, David Jones, Keiran Raine, Calli Latimer, Claudio R Santos, Mahrokh Nohadani, Aron C Eklund, Bradley Spencer-Dene, Graham Clark, Lisa Pickering, Gordon Stamp, Martin Gore, Zoltan Szallasi, Julian Downward, P Andrew Futreal, and Charles Swanton. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.*, 366(10):883–892, 2012.

[145] Lucy R Yates, Moritz Gerstung, Stian Knappskog, Christine Desmedt, Gunes Gundem, Peter Van Loo, Turid Aas, Ludmil B Alexandrov, Denis Larsimont, Helen Davies, Yilong Li, Young Seok Ju, Manasa Ramakrishna, Hans Kristian Haugland, Peer Kaare Lilleng, Serena Nik-Zainal, Stuart McLaren, Adam Butler, Sancha Martin, Dominic Glodzik, Andrew Menzies, Keiran Raine,

Jonathan Hinton, David Jones, Laura J Mudie, Bing Jiang, Delphine Vincent, April Greene-Colozzi, Pierre-Yves Adnet, Aquila Fatima, Marion Maetens, Michail Ignatiadis, Michael R Stratton, Christos Sotiriou, Andrea L Richardson, Per Eystein Lønning, David C Wedge, and Peter J Campbell. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nature Medicine*, 21(7):751–9, 2015.

[146] Samra Turajlic and Charles Swanton. Metastasis as an evolutionary process. *Science (New York, N.Y.)*, 352(6282):169–175, 4 2016.

[147] Zofia Piotrowska, Matthew J. Niederst, Chris A. Karlovich, Heather A. Wakelee, Joel W. Neal, Mari Mino-Kenudson, Linnea Fulton, Aaron N. Hata, Elizabeth L. Lockerman, Anuj Kalsy, Subba Digumarthy, Alona Muzikansky, Mitch Raponi, Angel R. Garcia, Hillary E. Mulvey, Melissa K. Parks, Richard H. DiCecca, Dora Dias-Santagata, A. John Iafrate, Alice T. Shaw, Andrew R. Allen, Jeffrey A. Engelman, and Lecia V. Sequist. Heterogeneity Underlies the Emergence of EGFRT790 Wild-Type Clones Following Treatment of T790M-Positive Cancers with a Third-Generation EGFR Inhibitor. *Cancer Discovery*, 5(7):37–54, 7 2015.

[148] Yung Bin Kuo, Jinn Shiun Chen, Chung Wei Fand, Yi Shuan Li, and Err Cheng Chan. Comparison of KRAS mutation analysis of primary tumors and matched circulating cell-free DNA in plasmas of patients with colorectal cancer. *Clinica Chimica Acta*, 433:284–289, 2014.

[149] Jianjun Zhang, Junya Fujimoto, Jianhua Zhang, David C Wedge, Xingzhi Song, Jiexin Zhang, Sahil Seth, Chi Wan Chow, Yu Cao, Curtis Gumbs, Kathryn A Gold, Neda Kalhor, Latasha Little, Harshad Mahadeshwar, Cesar Moran, Alexei Protopopov, Huandong Sun, Jiabin Tang, Xifeng Wu, Yuanqing Ye, William N William, Jack Jack Lee, John V Heymach, Waun Ki Hong, Stephen Swisher, Ignacio I Wistuba, and P Andrew Futreal. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science*, 346(6206):256–259, 10 2014.

[150] Elza C de Bruin, Nicholas McGranahan, Richard Mitter, Max Salm, David C Wedge, Lucy Yates, Mariam Jamal-Hanjani, Seema Shafi, Nirupa Murugaesu, Andrew J Rowan, Eva Grönroos, Madiha A Muhammad, Stuart Horswell, Marco Gerlinger, Ignacio Varela, David Jones, John Marshall, Thierry Voet, Peter Van Loo, Doris M Rassl, Robert C Rintoul, Sam M Janes, Siow-Ming Lee, Martin Forster, Tanya Ahmad, David Lawrence, Mary Falzon, Arrigo Capitanio, Timothy T Harkins, Clarence C Lee, Warren Tom, Enock Teefe, Shann-Ching Chen, Sharmin Begum, Adam Rabinowitz, Benjamin Phillimore, Bradley Spencer-Dene, Gordon Stamp, Zoltan Szallasi, Nik Matthews, Aengus Stewart, Peter Campbell, and Charles Swanton. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science (New York, N.Y.)*, 346(6206):251–256, 2014.

[151] Giulia Siravegna, Benedetta Mussolin, Michela Buscarino, Giorgio Corti, Andrea Cassingena, Giovanni Crisafulli, Agostino Ponzetti, Chiara Cremolini, Alessio Amatu, Calogero Lauricella, Simona Lamba, Sebastijan Hobor, Antonio

Avallone, Emanuele Valtorta, Giuseppe Rospo, Enzo Medico, Valentina Motta, Carlotta Antoniotti, Fabiana Tatangelo, Beatriz Bellosillo, Silvio Veronese, Alfredo Budillon, Clara Montagut, Patrizia Racca, Silvia Marsoni, Alfredo Falcone, Ryan B Corcoran, Federica Di Nicolantonio, Fotios Loupakis, Salvatore Siena, Andrea Sartore-Bianchi, and Alberto Bardelli. Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nature Medicine*, 21(7):795–801, 2015.

[152] Martin Reck, Koichi Hagiwara, Baohui Han, Sergei Tjulandin, Christian Grohé, Takashi Yokoi, Alessandro Morabito, Silvia Novello, Edurne Arriola, Olivier Molinier, Rose McCormack, Marianne Ratcliffe, and Nicola Normanno. Circulating Free Tumor-derived DNA (ctDNA) Determination of EGFR Mutation Status in European and Japanese Patients with Advanced NSCLC: the ASSESS Study. *Journal of Thoracic Oncology*, 11(10):1682–1689, 2016.

[153] Jean-Yves Douillard, Gyula Ostoros, Manuel Cobo, Tudor Ciuleanu, Rebecca Cole, Gael McWalter, Jill Walker, Simon Dearden, Alan Webster, Tsveta Milenkova, and Rose McCormack. Gefitinib Treatment in EGFR Mutated Caucasian NSCLC. *Journal of Thoracic Oncology*, 9(9):1345–1353, 2014.

[154] Xin Qian, Jia Liu, Yuhui Sun, Meifang Wang, Huaiding Lei, Guoshi Luo, Xianjun Liu, Chang Xiong, Dan Liu, Jie Liu, and Yijun Tang. Circulating cell-free DNA has a high degree of specificity to detect exon 19 deletions and the single-point substitution mutation L858R in non-small cell lung cancer. *Oncotarget*, 7(20):29154–29165, 5 2016.

[155] J. Y. Douillard, G Ostoros, M Cobo, T Ciuleanu, R McCormack, A Webster, and T Milenkova. First-line gefitinib in Caucasian EGFR mutation-positive NSCLC patients: A phase-IV, open-label, single-arm study. *British Journal of Cancer*, 110(1):55–62, 2014.

[156] Tony S. Mok, Yi-Long Wu, Myung-Ju Ahn, Marina C. Garassino, Hye R. Kim, Suresh S. Ramalingam, Frances A. Shepherd, Yong He, Hiroaki Akamatsu, Willemijn S M E Theelen, Chee K. Lee, Martin Sebastian, Alison Templeton, Helen Mann, Marcelo Marotti, Serban Ghiorghiu, Vassiliki A. Papadimitrakopoulou, and AURA3 Investigators. Osimertinib or Platinum-Pemetrexed in EGFR T790M-Positive Lung Cancer. *The New England Journal of Medicine*, 376(7):629–640, 2017.

[157] Kenneth S. Thress, Roz Brant, T. Hedley Carr, Simon Dearden, Suzanne Jenkins, Helen Brown, Tracey Hammett, Mireille Cantarini, and J. Carl Barrett. EGFR mutation detection in ctDNA from NSCLC patient plasma: A cross-platform comparison of leading technologies to support the clinical development of AZD9291. *Lung Cancer*, 90(3):509–515, 2015.

[158] Sumitra Mohan, Ellen Heitzer, Peter Ulz, Ingrid Lafer, Sigurd Lax, Martina Auer, Martin Pichler, Armin Gerger, Florian Eisner, Gerald Hoefler, Thomas Bauernhofer, Jochen B. Geigl, and Michael R. Speicher. Changes in Colorectal Carcinoma Genomes under Anti-EGFR Therapy Identified by Whole-Genome Plasma DNA Sequencing. *PLoS Genetics*, 10(3):e1004271, 2014.

[159] Isaac Kinde, Nickolas Papadopoulos, Kenneth W. Kinzler, and Bert Vogelstein. FAST-SeqS: A simple and efficient method for the detection of aneuploidy by massively parallel sequencing. *PLoS ONE*, 7(7):e41162, 2012.

[160] Diana Abdueva, Darya Chudova, Richard Lanman, Kimberly Banks, Rebecca Nagy, Stefanie Mortimer, AmirAli Talasaz, Helmy Eltoukhy, and Razelle Kurzrock. Abstract 11541: Detection, frequency and actionability of recurrent copy number gains detected by non-invasive liquid biopsy of 3,942 lung and breast cancer samples. *Journal of Clinical Oncology*, 34(Supplementary), 2016.

[161] Jacob J. Chabon, Andrew D. Simmons, Alexander F. Lovejoy, Mohammad S. Esfahani, Aaron M. Newman, Henry J. Haringsma, David M. Kurtz, Henning Stehr, Florian Scherer, Chris A. Karlovich, Thomas C. Harding, Kathleen A. Durkin, Gregory A. Otterson, W. Thomas Purcell, D. Ross Camidge, Jonathan W. Goldman, Lecia V. Sequist, Zofia Piotrowska, Heather A. Wakelee, Joel W. Neal, Ash A. Alizadeh, and Maximilian Diehn. Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients. *Nature Communications*, 7:11815, 2016.

[162] Peiyong Jiang, Carol W. M. Chan, K. C. Allen Chan, Suk Hang Cheng, John Wong, Vincent Wai-Sun Wong, Grace L. H. Wong, Stephen L. Chan, Tony S. K. Mok, Henry L. Y. Chan, Paul B. S. Lai, Rossa W. K. Chiu, and Y. M. Dennis Lo. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proceedings of the National Academy of Sciences*, 112(11):1317–25, 2015.

[163] Alessandro Romanel, Delila Gasi Tandefelt, Vincenza Conteduca, Anuradha Jayaram, Nicola Casiraghi, Daniel Wetterskog, Samanta Salvi, Dino Amadori, Zafeiris Zafeiriou, Pasquale Rescigno, Diletta Bianchini, Giorgia Gurioli, Valentina Casadio, Suzanne Carreira, Jane Goodall, Anna Wingate, Roberta Ferraldeschi, Nina Tunariu, Penny Flohr, Ugo De Giorgi, Johann S De Bono, Francesca Demichelis, and Gerhardt Attard. Plasma AR and abiraterone-resistant prostate cancer. *Science Transl. Med.*, 7(312):10–312, 2015.

[164] Cloud P. Paweletz, Adrian G. Sacher, Chris K. Raymond, Ryan S. Alden, Allison O'Connell, Stacy L. MacH, Yanan Kuang, Leena Gandhi, Paul Kirschmeier, Jessie M. English, Lee P. Lim, Pasi A. Janne, and Geoffrey R. Oxnard. Bias-corrected targeted next-generation sequencing for rapid, multiplexed detection of actionable alterations in cell-free DNA from advanced lung cancer patients. *Clinical Cancer Research*, 22(4):915–922, 2016.

[165] Office for National Statistics. Cancer survival by stage at diagnosis for England (experimental statistics): Adults diagnosed 2012 , 2013 and 2014 and followed up to 2015, 2016.

[166] Emmanuelle Gormally, Paolo Vineis, Giuseppe Matullo, Fabrizio Veglia, Elodie Caboux, Emilie Le Roux, Marco Peluso, Seymour Garte, Simonetta Guarrera, Armelle Munnia, Luisa Airoldi, Herman Autrup, Christian Malaveille, Alison Dunning, Kim Overvad, Anne Tjønneland, Eiliv Lund, Françoise Clavel-Chapelon, Heiner Boeing, Antonia Trichopoulou, Domenico Palli, Vittorio

Krogh, Rosario Tumino, Salvatore Panico, H. Bas Bueno-De-Mesquita, Petra H. Peeters, Guillem Pera, Carmen Martinez, Miren Dorronsoro, Aurelio Barricarte, Carmen Navarro, José Ramón Quirós, Göran Hallmans, Nicholas E. Day, Timothy J. Key, Rodolfo Saracci, Rudolf Kaaks, Elio Riboli, and Pierre Hainaut. TP53 and KRAS2 mutations in plasma DNA of healthy subjects and subsequent cancer occurrence: A prospective study. *Cancer Research*, 66(13):6871–6876, 2006.

[167] Frédéric Amant, Magali Verheecke, Iwona Wlodarska, Luc Dehaspe, Paul Brady, Nathalie Brison, Kris Van Den Bogaert, Daan Dierickx, Vincent Vandecaveye, Thomas Tousseyn, Philippe Moerman, Adriaan Vanderstichele, Ignace Vergote, Patrick Neven, Patrick Berteloot, Katrien Putseys, Lode Danneels, Peter Vandenberghe, Eric Legius, and Joris Robert Vermeesch. Presymptomatic Identification of Cancers in Pregnant Women During Noninvasive Prenatal Testing. *JAMA Oncology*, 1(6):814, 2015.

[168] Diana W. Bianchi, Darya Chudova, Amy J. Sehnert, Sucheta Bhatt, Kathryn Murray, Tracy L. Prosen, Judy E. Garber, Louise Wilkins-Haug, Neeta L. Vora, Stephen Warsof, James Goldberg, Tina Ziainia, and Meredith Halks-Miller. Noninvasive Prenatal Testing and Incidental Detection of Occult Maternal Malignancies. *JAMA*, 314(2):162, 2015.

[169] Julia A Beaver, Danijela Jelovac, Sasidharan Balukrishna, Rory L Cochran, Sarah Croessmann, Daniel J Zabransky, Hong Yuen Wong, Patricia Valda Toro, Justin Cidado, Brian G Blair, David Chu, Timothy Burns, Michaela J Higgins, Vered Stearns, Lisa Jacobs, Mehran Habibi, Julie Lange, Paula J Hurley, Josh Lauring, Dustin A VanDenBerg, Jill Kessler, Stacie Jeter, Michael L Samuels, Dianna Maar, Leslie Cope, Ashley Cimino-Mathews, Pedram Argani, Antonio C Wolff, and Ben Ho Park. Detection of cancer DNA in plasma of patients with early-stage breast cancer. *Clinical Cancer Research*, 20(10):2643–50, 2014.

[170] Paul A. Cohen, Nicola Flowers, Stephen Tong, Natalie Hannan, Mark D. Pertile, and Lisa Hui. Abnormal plasma DNA profiles in early ovarian cancer using a non-invasive prenatal testing platform: implications for cancer screening. *BMC Medicine*, 14(1):126, 2016.

[171] Yao Wang, Mai Chen, Nan Xiao, and Hui Liu. Evaluation and comparison of in vitro degradation kinetics of DNA in serum, urine and saliva: A qualitative Study. *Gene*, 590(1):142–148, 2016.

[172] Max Schreuer, Geert Meersseman, Sari Van Den Herrewegen, Yanina Jansen, Ines Chevolet, Ambre Bott, Sofie Wilgenhof, Teofila Seremet, Bart Jacobs, Ronald Buyl, Geert Maertens, and Bart Neyns. Quantitative assessment of BRAF V600 mutant circulating cell-free tumor DNA as a tool for therapeutic monitoring in metastatic melanoma patients treated with BRAF/MEK inhibitors. *Journal of Translational Medicine*, 14(1):95, 2016.

[173] Antonio Marchetti, John F Palma, Lara Felicioni, Tommaso M De Pas, Rita Chiari, Maela Del Grammastro, Giampaolo Filice, Vienna Ludovini, Alba A Brandes, Antonio Chella, Francesco Malorgio, Flavio Guglielmi, Michele

De Tursi, Armando Santoro, Lucio Crinò, and Fiamma Buttitta. Early prediction of response to tyrosine kinase inhibitors by quantification of EGFR mutations in plasma of NSCLC patients. *Journal of Thoracic Oncology*, 10(10):1437–1443, 2015.

[174] Liqiang Xi, Trinh Pham, Eden C Payabyab, Richard M Sherry, Steven A Rosenberg, and Mark Raffeld. Circulating Tumor DNA as an Early Indicator of Response to T-Cell Transfer Immunotherapy in Metastatic Melanoma. *Clinical Cancer Research*, 22(22):5480–5486, 8 2016.

[175] J. Tie, I. Kinde, Y. Wang, H. L. Wong, J. Roebert, M. Christie, M. Tacey, R. Wong, M. Singh, C. S. Karapetis, J. Desai, B. Tran, R. L. Strausberg, L. A. Diaz, N. Papadopoulos, K. W. Kinzler, B. Vogelstein, and P. Gibbs. Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer. *Annals of Oncology*, 26(8):1715–1722, 2015.

[176] Yi Xin, Lu Xiaoxing, Zhao Meiru, Guan Yanfang, Liu Tao, and Yang Ling. Patent CN105087789A - Method for detecting BCR and TCR immune repertoire in blood plasma cfDNA, 11 2015.

[177] Sandra Misale, Rona Yaeger, Sebastijan Hobor, Elisa Scala, David Liska, Emanuele Valtorta, Roberta Schiavo, Michela Buscarino, Giulia Siravegna, Katia Bencardino, Andrea Cercek, Chin-tung Chen, Silvio Veronese, Carlo Zanon, Andrea Sartore-bianchi, Marcello Gambacorta, Margherita Gallicchio, Efsevia Vakiani, Valentina Boscaro, Enzo Medico, Martin Weiser, and Salvatore Siena. Emergence of KRAS mutations and acquired resistance to anti EGFR therapy in colorectal cancer. *Nature*, 486(7404):532–536, 2012.

[178] Luis A. Diaz, Richard T Williams, Jian Wu, Isaac Kinde, J Randolph Hecht, Jordan Berlin, Benjamin Allen, Ivana Bozic, Johannes G Reiter, Martin A Nowak, Kenneth W Kinzler, Kelly S Oliner, and Bert Vogelstein. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature*, 486(7404):537–540, 6 2012.

[179] Marius Ilié and Paul Hofman. Pros: Can tissue biopsy be replaced by liquid biopsy? *Translational Lung Cancer Research*, 5(4):420–423, 2016.

[180] Mariangela Russo, Giulia Siravegna, Lawrence S Blaszkowsky, Giorgio Corti, Giovanni Crisafulli, Leanne G Ahronian, Benedetta Mussolin, Eunice L Kwak, Michela Buscarino, Luca Lazzari, Emanuele Valtorta, Mauro Truini, Nicholas A Jessop, Hayley E Robinson, Theodore S Hong, Mari Mino-Kenudson, Federica Di Nicolantonio, Ashraf Thabet, Andrea Sartore-Bianchi, Salvatore Siena, A John Iafrate, Alberto Bardelli, and Ryan B Corcoran. Tumor heterogeneity and lesion-specific response to targeted therapy in colorectal cancer. *Cancer Discovery*, 6(2):147–53, 2015.

[181] M. P. Morelli, M. J. Overman, A. Dasari, S. M. a. Kazmi, T. Mazard, E. Vilar, V. K. Morris, M. S. Lee, D. Herron, C. Eng, J. Morris, B. K. Kee, F. Janku, F. L. Deaton, C. Garrett, D. Maru, F. Diehl, P. Angenendt, and S. Kopetz. Characterizing the patterns of clonal selection in circulating tumor DNA from

patients with colorectal cancer refractory to anti-EGFR treatment. *Annals of Oncology*, 26(4):731–736, 2015.

[182] G Gremel, R J Lee, M R Girotti, A K Mandal, S Valpione, G Garner, M Ayub, S Wood, D G Rothwell, A Fusi, A Wallace, G Brady, C Dive, N Dhomen, P Lorigan, R Marais, and Richard Marais. Distinct sub-clonal tumour responses to therapy revealed by circulating cell-free DNA. *Ann. Oncol.*, 27(10):1959–1965, 2016.

[183] Peter Ulz, Jelena Belic, Ricarda Graf, Martina Auer, Ingrid Lafer, Katja Fischereder, Gerald Webersinke, Karl Pummer, Herbert Augustin, Martin Pichler, Gerald Hoefler, Thomas Bauernhofer, Jochen B Geigl, Ellen Heitzer, and Michael R Speicher. Whole-genome plasma sequencing reveals focal amplifications as a driving force in metastatic prostate cancer. *Nature Communications*, 7:12008, 2016.

[184] Mariangela Russo, Sandra Misale, Ge Wei, Giulia Siravegna, Giovanni Crisafulli, Luca Lazzari, Giorgio Corti, Giuseppe Rospo, Luca Novara, Benedetta Mussolin, Alice Bartolini, Nicholas Cam, Roopal Patel, Shunqi Yan, Robert Shoemaker, Robert Wild, Federica di Nicolantonio, Andrea Sartore-Bianchi, Gang Li, Salvatore Siena, and Alberto Bardelli. Acquired resistance to the TRK inhibitor entrectinib in colorectal cancer. *Cancer Discovery*, 6(1):36–44, 2016.

[185] Jeanne Tie, Yuxuan Wang, Cristian Tomasetti, Lu Li, Simeon Springer, Isaac Kinde, Natalie Silliman, Mark Tacey, Hui-li Wong, Michael Christie, Suzanne Kosmider, Iain Skinner, Rachel Wong, Malcolm Steel, Ben Tran, Jayesh Desai, Ian Jones, Andrew Haydon, Theresa Hayes, Tim J Price, Robert L Strausberg, Luis A Diaz Jr, Nickolas Papadopoulos, Kenneth W Kinzler, Bert Vogelstein, and Peter Gibbs. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Science Transl. Med.*, 8(346):92–346, 2016.

[186] Isaac Garcia-Murillas, Gaia Schiavon, Britta Weigelt, Charlotte Ng, Sarah Hrebien, Rosalind J Cutts, Maggie Cheang, Peter Osin, Ashutosh Nerurkar, Iwanka Kozarewa, Javier Armisen Garrido, Mitch Dowsett, J. S. Reis-Filho, Ian E Smith, and Nicholas C Turner. Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Science Translational Medicine*, 7(302):133–302, 2015.

[187] Thomas Reinert, Lone V Schøler, Rune Thomsen, Heidi Tobiasen, Søren Vang, Iver Nordentoft, Philippe Lamy, Anne-Sofie Kannerup, Frank V Mortensen, Katrine Stribolt, Stephen Hamilton-Dutoit, Hans J Nielsen, Søren Laurberg, Niels Pallisgaard, Jakob S Pedersen, Torben F Ørntoft, and Claus L Andersen. Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery. *Gut*, 65(4):625–34, 2015.

[188] Rebecca J Leary, Isaac Kinde, Frank Diehl, Kerstin Schmidt, Chris Clouser, Cisilya Duncan, Alena Antipova, Clarence Lee, Kevin McKernan, Francisco M De La Vega, Kenneth W Kinzler, Bert Vogelstein, Luis a Diaz, and Victor E Velculescu. Development of personalized tumor biomarkers using massively parallel sequencing. *Science Transl. Med.*, 2:20ra14, 2010.

[189] David J Mcbride, Arto K Orpana, Christos Sotiriou, Heikki Joensuu, J Philip, Laura J Mudie, Eija Hämälaïnen, Lucy A Stebbings, Leif C Andersson, Adrienne M Flanagan, Virginie Durbecq, Michail Ignatiadis, Olli Kallioniemi, A Heckman, Kari Alitalo, Henrik Edgren, P Andrew Futreal, and Michael R Stratton. Use of Cancer-Specific Genomic Rearrangements to Quantify Disease Burden in Plasma from Patients with Solid Tumors. *Genes Chromosomes Cancer*, 49(11):1062–1069, 2011.

[190] R. Guerrero-Preston, B. L. Valle, A. Jedlicka, N. Turaga, O. Folawiyo, F. Pirini, F. Lawson, A. Vergura, M. G. Noordhuis, A. Dziedzic, G. Perez, M. Renehan, C. Guerrero-Diaz, E. De Jesus-Rodriguez, T. Diaz-Montes, J. Rodriguez-Orengo, K. Mendez, J. Romaguera, B. J. Trock, L. Florea, and D. Sidransky. Molecular Triage of Premalignant Lesions in Liquid-Based Cervical Cytology and Circulating Cell-Free DNA from Urine, Using a Panel of Methylated Human Papilloma Virus and Host Genes. *Cancer Prevention Research*, 9(12):915–924, 2016.

[191] K C Allen Chan, John K S Woo, Ann King, Benny C Y Zee, W K Jacky Lam, Stephen L Chan, Sam W I Chu, Constance Mak, Irene O L Tse, Samantha Y M Leung, Gloria Chan, Edwin P Hui, Brigette B Y Ma, Rossa W K Chiu, Sing-Fai Leung, Andrew C van Hasselt, Anthony T C Chan, and Y M Dennis Lo. Analysis of Plasma Epstein-Barr Virus DNA to Screen for Nasopharyngeal Cancer. *The New England Journal of Medicine*, 377(6):513–522, 8 2017.

[192] K M Patel, K E van der Vos, C G Smith, F Mouliere, D Tsui, J Morris, D Chandrananda, F Marass, D van den Broek, D E Neal, V J Gnanapragasam, T Forshew, B W van Rhijn, C E Massie, N Rosenfeld, and M S van der Heijden. Association Of Plasma And Urinary Mutant DNA With Clinical Outcomes In Muscle Invasive Bladder Cancer. *Scientific Reports*, 7(1):5554, 2017.

[193] Isaac Kinde, Chetan Bettegowda, Yuxuan Wang, Jian Wu, Nishant Agrawal, Ie-Ming Shih, Robert Kurman, Fanny Dao, Douglas A Levine, Robert Giuntoli, Richard Roden, James R Eshleman, Jesus Paula Carvalho, Suely Kazue Nagahashi Marie, Nickolas Papadopoulos, Kenneth W Kinzler, Bert Vogelstein, and Luis A Diaz. Evaluation of DNA from the Papanicolaou test to detect ovarian and endometrial cancers. *Science Translational Medicine*, 5(167):167ra4, 2013.

[194] Navya Nair, Olga Camacho-Vanegas, Dmitry Rykunov, Matthew Dashkoff, Sandra Catalina Camacho, Cassie A. Schumacher, Jonathan C. Irish, Timothy T. Harkins, Elijah Freeman, Isaac Garcia, Elena Pereira, Sviatoslav Kendall, Rachel Belfer, Tamara Kalir, Robert Sebra, Boris Reva, Peter Dottino, and John A. Martignetti. Genomic Analysis of Uterine Lavage Fluid Detects Early Endometrial Cancers and Reveals a Prevalent Landscape of Driver Mutations in Women without Histopathologic Evidence of Cancer: A Prospective Cross-Sectional Study. *PLoS Medicine*, 13(12):e1002206, 2016.

[195] Caryn S Ross-Innes, Hamza Chettouh, Achilleas Achilleos, Nuria Galeano-Dalmau, Irene Debiram-Beecham, Shona MacRae, Petros Fessas, Elaine Walker, Sibu Varghese, Theodore Evan, Pierre S Lao-Sirieix, Maria O'Donovan, Shalini Malhotra, Marco Novelli, Babett Disep, Phillip V Kaye, Laurence B Lovat,

Rehan Haidry, Michael Griffin, Krish Ragunath, Pradeep Bhandari, Adam Haycock, Danielle Morris, Stephen Attwood, Anjan Dhar, Colin Rees, Matt D Rutter, Richard Ostler, Benoit Aigret, Peter D Sasieni, and Rebecca C Fitzgerald. Risk stratification of Barrett's oesophagus using a non-endoscopic sampling method coupled with a biomarker panel: a cohort study. *The Lancet Gastroenterology & Hepatology*, 2(1):23–31, 12 2016.

[196] Tobias M Gorges, Nicole Penkalla, Thomas Schalk, Simon A Joosse, Sabine Riethdorf, Johannes Tucholski, Klaus Lücke, Harriet Wikman, Stephen Jackson, Nora Brychta, Oliver von Ahsen, Christian Schumann, Thomas Krahn, and Klaus Pantel. Enumeration and Molecular Characterization of Tumor Cells in Lung Cancer Patients Using a Novel In Vivo Device for Capturing Circulating Tumor Cells. *Clinical Cancer Research*, 22(9):2197–206, 2016.

[197] Louise Aigrain, Yong Gu, and Michael A Quail. Quantitation of next generation sequencing library preparation protocol efficiencies using droplet digital PCR assays - a systematic comparison of DNA library preparation kits for Illumina sequencing. *BMC Genomics*, 17:458, 6 2016.

[198] Florent Mouliere, Bruno Robert, Erika Arnau Peyrotte, Maguy Del Rio, Marc Ychou, Franck Molina, Celine Gongora, and Alain R. Thierry. High Fragmentation Characterizes Tumour-Derived Circulating DNA. *PLoS ONE*, 6(9):e23418, 2011.

[199] Florent Mouliere, Anna M. Piskorz, Dineika Chandrananda, Elizabeth Moore, James Morris, Christopher G. Smith, Teodora Goranova, Katrin Heider, Richard Mair, Anna Supernat, Ioannis Gounaris, Susana Ros, Jonathan C. M. Wan, Mercedes Jimenez-Linan, Davina Gale, Kevin Brindle, Charles E. Massie, Christine A. Parkinson, James D. Brenton, and Nitzan Rosenfeld. Selecting Short DNA Fragments In Plasma Improves Detection Of Circulating Tumour DNA. *bioRxiv*, 1 2017.

[200] E. Pienaar, M. Theron, M. Nelson, and H. J. Viljoen. A quantitative model of error accumulation during PCR amplification. *Computational Biology and Chemistry*, 30(2):102–111, 2006.

[201] Melanie Schirmer, Rosalinda D'Amore, Umer Z Ijaz, Neil Hall, and Christopher Quince. Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics*, 17(1):125, 3 2016.

[202] Franziska Pfeiffer, Carsten Gröber, Michael Blank, Kristian Händler, Marc Beyer, Joachim L. Schultze, and Günter Mayer. Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Scientific Reports*, 8(1):1–14, 2018.

[203] Jillian Phallen, Mark Sausen, Vilmos Adleff, Alessandro Leal, Carolyn Hruban, James White, Valsamo Anagnostou, Jacob Fiksel, Stephen Cristiano, Eniko Papp, Savannah Speir, Thomas Reinert, Mai-Britt Worm Orntoft, Brian D Woodward, Derek Murphy, Sonya Parpart-Li, David Riley, Monica Nesselbush, Naomi Sengamalay, Andrew Georgiadis, Qing Kay Li, Mogens Rørbæk Madsen,

Frank Viborg Mortensen, Joost Huiskens, Cornelis Punt, Nicole van Grieken, Remond Fijneman, Gerrit Meijer, Hatim Husain, Robert B Scharpf, Luis A Diaz, Siân Jones, Sam Angiuoli, Torben Ørntoft, Hans Jørgen Nielsen, Claus Lindbjerg Andersen, and Victor E Velculescu. Direct detection of early-stage cancers using circulating tumor DNA. *Science Translational Medicine*, 9(403), 8 2017.

[204] K C Allen Chan, Peiyong Jiang, Kun Sun, Yvonne K Y Cheng, Yu K Tong, Suk Hang Cheng, Ada I C Wong, Irena Hudecova, Tak Y Leung, Rossa W K Chiu, and Yuk Ming Dennis Lo. Second generation noninvasive fetal genome analysis reveals de novo mutations, single-base parental inheritance, and preferred DNA ends. *Proceedings of the National Academy of Sciences*, 113(50):E8159–E8168, 12 2016.

[205] Anthony M Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120, 8 2014.

[206] Nitzan Rosenfeld, Tim Forshew, Francesco Marass, and Muhammed Murtaza. Patent WO2016009224 A1 - A method for detecting a genetic variant, 2016.

[207] Margaret L. Hoang, Isaac Kinde, Cristian Tomasetti, K. Wyatt McMahon, Thomas A. Rosenquist, Arthur P. Grollman, Kenneth W. Kinzler, Bert Vogelstein, and Nickolas Papadopoulos. Genome-wide quantification of rare somatic mutations in normal human tissues using massively parallel sequencing. *Proceedings of the National Academy of Sciences*, 113(35):9846–9851, 8 2016.

[208] H. Christina Fan, Yair J. Blumenfeld, Usha Chitkara, Louanne Hudgins, and Stephen R. Quake. Analysis of the size distributions of fetal and maternal cell-free DNA by paired-end sequencing. *Clinical Chemistry*, 56(8):1279–1286, 2010.

[209] I. Martincorena, A. Roshan, M. Gerstung, P. Ellis, P. Van Loo, S. McLaren, D. C. Wedge, A. Fullam, L. B. Alexandrov, J. M. Tubio, L. Stebbings, A. Menzies, S. Widaa, M. R. Stratton, P. H. Jones, and P. J. Campbell. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science (New York, N.Y.)*, 348(6237):880–886, 2015.

[210] Giulio Genovese, Anna K Kähler, Robert E Handsaker, Johan Lindberg, Samuel A Rose, Samuel F Bakhoum, Kimberly Chambert, Eran Mick, Benjamin M Neale, Menachem Fromer, Shaun M Purcell, Oscar Svantesson, Mikael Landén, Martin Höglund, Sören Lehmann, Stacey B Gabriel, Jennifer L Moran, Eric S Lander, Patrick F Sullivan, Pamela Sklar, Henrik Grönberg, Christina M Hultman, and Steven A McCarroll. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *The New England Journal of Medicine*, 371(26):2477–87, 2014.

[211] European Medicines Agency. Guideline on good pharmacogenomic practice, 2016.

[212] Aarhus University Hospital. NCT02284633 - Blood Sample Monitoring of Patients With EGFR Mutated Lung Cancer, 2015.

[213] The Institute of Cancer Research Clinical Trials & Statistics Unit. plasmaMATCH: A clinical trial aiming to assess the safety and activity of targeted treatments in patients with advanced breast cancer where the targetable mutation is identified through circulating tumour DNA screening, 2016.

[214] Shalin H. Naik, Ton N. Schumacher, and Leila Perie. Cellular barcoding: A technical appraisal. *Experimental Hematology*, 42(8):598–608, 2014.

[215] Safia El Messaoudi, Florent Mouliere, Stanislas Du Manoir, Caroline Bascoul-Mollevi, Brigitte Gillet, Michelle Nouaille, Catherine Fiess, Evelyne Crapez, Frederic Bibeau, Charles Theillet, Thibault Mazard, Denis Pezet, Muriel Mathonnet, Marc Ychou, and Alain R Thierry. Circulating DNA as a Strong Multimarker Prognostic Tool for Metastatic Colorectal Cancer Patient Management Care. *Clinical Cancer Research*, 22(12):3067–3077, 6 2016.

[216] Karen Lise Garm Spindler, Niels Pallisgaard, Rikke Fredslund Andersen, and Anders Jakobsen. Changes in mutational status during third-line treatment for metastatic colorectal cancer - Results of consecutive measurement of cell free DNA, KRAS and BRAF in the plasma. *International Journal of Cancer*, 135(9):2215–2222, 2014.

[217] Heidi Schwarzenbach, Dave S B Hoon, and Klaus Pantel. Cell-free nucleic acids as biomarkers in cancer patients. *Nat. Rev. Cancer*, 11(6):426–437, 5 2011.

[218] R J Sullivan, V J O'Neill, D Enderle, M Valentino, T Koestler, S Blackmon, A K Krug, K Brinkmann, A Spiel, S Bentink, R Mueller, J Emenegger, M Noerholm, J Skog, C Berking, and K Flaherty. Plasma-Based Monitoring of BRAF Mutations During Therapy for Malignant Melanoma Using Combined Exosomal RNA and Cell-Free DNA Analysis. *Journal of Clinical Oncology*, 33(15_suppl):9017, 2015.

[219] Huilin Shao, Jaehoon Chung, Kyungheon Lee, Leonora Balaj, Changwook Min, Bob S. Carter, Fred H. Hochberg, Xandra O. Breakefield, Hakho Lee, and Ralph Weissleder. Chip-based analysis of exosomal mRNA mediating drug resistance in glioblastoma. *Nature Communications*, 6(May):6999, 2015.

[220] Catherine Alix-Panabières and Klaus Pantel. Clinical Applications of Circulating Tumor Cells and Circulating Tumor DNA as Liquid Biopsy. *Cancer Discovery*, 6(5):479–491, 2016.

[221] Myron G G. Best, Nik Sol, Irsan Kooi, Jihane Tannous, Bart A A. Westerman, François Rustenburg, Pepijn Schellen, Heleen Verschueren, Edward Post, Jan Koster, Bauke Ylstra, Najim Ameziane, Josephine Dorsman, Egbert F F. Smit, Henk M M. Verheul, David P P. Noske, Jaap C C. Reijneveld, R Jonas A Jonas A Nilsson, Bakhos A A. Tannous, Pieter Wesseling, and Thomas Wurdinger. RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics. *Cancer Cell*, 28(5):666–676, 11 2015.

[222] Emmanuelle Gormally, Elodie Caboux, Paolo Vineis, and Pierre Hainaut. Circulating free DNA in plasma or serum as biomarker of carcinogenesis: Practical aspects and biological significance, 2007.

[223] Sarah Webb. The cancer bloodhounds. *Nature Biotechnology*, 34(11):1090–1094, 2016.

[224] Elodie Long-Mira, Kevin Washetine, and Paul Hofman. Sense and nonsense in the process of accreditation of a pathology laboratory. *Virchows Archiv*, 468(1):43–49, 2016.

[225] Linea Melchior, Morten Grauslund, Beatriz Bellosillo, Clara Montagut, Erica Torres, Ester Moragen, Isabel Micalessi, Johan Frans, Veerle Noten, Claire Bourgain, Renske Vriesema, Robert van der Geize, Kristof Cokelaere, Nancy Vercooren, Katrien Crul, Thomas R??diger, Diana Buchm??ller, Martin Reijans, and Caroline Jans. Multi-center evaluation of the novel fully-automated PCR-based Idylla BRAF Mutation Test on formalin-fixed paraffin-embedded tissue of malignant melanoma. *Experimental and Molecular Pathology*, 99(3):485–491, 2015.

[226] Filip Janku, Helen J Huang, Bart Claes, Gerald S Falchook, Siqing Fu, David Hong, Nishma M Ramzanali, Giovanni Nitti, Goran Cabrilo, Apostolia M Tsimberidou, Aung Naing, Sarina A Piha-Paul, Jennifer J Wheler, Daniel D Karp, Veronica R Holley, Ralph G Zinner, Vivek Subbiah, Rajyalakshmi Luthra, Scott Kopetz, Michael J Overman, Bryan K Kee, Sapna Patel, Benoit Devogelaere, Erwin Sablon, Geert Maertens, Gordon B Mills, Razelle Kurzrock, and Funda Meric-Bernstam. BRAF Mutation Testing in Cell-Free DNA from the Plasma of Patients with Advanced Cancers Using a Rapid, Automated Molecular Diagnostics System. *Molecular Cancer Therapeutics*, 15(6):1397–404, 2016.

[227] Alexis L. Norris, Rachael E. Workman, Yunfan Fan, James R. Eshleman, and Winston Timp. Nanopore sequencing detects structural variants in cancer. *Cancer Biology and Therapy*, 17(3):246–253, 2016.

[228] Joshua Quick, Nicholas J. Loman, Sophie Duraffour, Jared T. Simpson, Ettore Severi, Lauren Cowley, Joseph Akoi Bore, Raymond Koundouno, Gytis Dudas, Amy Mikhail, Nobila Ouédraogo, Babak Afrough, Amadou Bah, Jonathan H. J. Baum, Beate Becker-Ziaja, Jan Peter Boettcher, Mar Cabeza-Cabrerizo, Álvaro Camino-Sánchez, Lisa L. Carter, Juliane Doerrbecker, Theresa Enkirch, Isabel García-Dorival, Nicole Hetzelt, Julia Hinzmann, Tobias Holm, Liana Eleni Kafetzopoulou, Michel Koropogui, Abigael Kosgey, Eeva Kuisma, Christopher H. Logue, Antonio Mazzarelli, Sarah Meisel, Marc Mertens, Janine Michel, Didier Ngabo, Katja Nitzsche, Elisa Pallasch, Livia Victoria Patrono, Jasmine Portmann, Johanna Gabriella Repits, Natasha Y. Rickett, Andreas Sachse, Katrin Singethan, Inês Vitoriano, Rahel L. Yemanaberhan, Elsa G. Zekeng, Trina Racine, Alexander Bello, Amadou Alpha Sall, Ousmane Faye, Oumar Faye, N'Faly Magassouba, Cecelia V Williams, Victoria Amburgey, Linda Winona, Emily Davis, Jon Gerlach, Frank Washington, Vanessa Monteil, Marine Jourdain, Marion Bererd, Alimou Camara, Hermann Somlare, Abdoulaye Camara, Marianne Gerard, Guillaume Bado, Bernard Baillet, Déborah

Delaune, Koumpingnin Yacouba Nebie, Abdoulaye Diarra, Yacouba Savane, Raymond Bernard Pallawo, Giovanna Jaramillo Gutierrez, Natacha Milhano, Isabelle Roger, Christopher J. Williams, Facinet Yattara, Kuiama Lewandowski, James Taylor, Phillip Rachwal, Daniel J Turner, Georgios Pollakis, Julian A. Hiscox, David A. Matthews, Matthew K O'Shea, Andrew McD. Johnston, Duncan Wilson, Emma Hutley, Erasmus Smit, Antonino Di Caro, Roman Wölfel, Kilian Stoecker, Erna Fleischmann, Martin Gabriel, Simon A. Weller, Lamine Koivogui, Boubacar Diallo, Sakoba Keïta, Andrew Rambaut, Pierre Formenty, Stephan Günther, and Miles W. Carroll. Real-time, portable genome sequencing for Ebola surveillance. *Nature*, 530(7589):228–32, 2016.

[229] Shan Wei and Zev Williams. Rapid short-read sequencing and aneuploidy detection using minION nanopore technology. *Genetics*, 202(1):37–44, 2016.

[230] Camilla L.C. Ip, Matthew Loose, John R. Tyson, Mariateresa de Cesare, Bonnie L. Brown, Miten Jain, Richard M. Leggett, David A. Eccles, Vadim Zalunin, John M. Urban, Paolo Piazza, Rory J. Bowden, Benedict Paten, Solomon Mwaigwisya, Elizabeth M. Batty, Jared T. Simpson, Terrance P. Snutch, Ewan Birney, David Buck, Sara Goodwin, Hans J. Jansen, Justin O'Grady, and Hugh E. Olsen. MinION Analysis and Reference Consortium: Phase 1 data release and analysis. *F1000Research*, 4:1075, 2015.

[231] Matthew Loose, Sunir Malla, and Michael Stout. Real-time selective sequencing using nanopore technology. *Nature Methods*, 13(9):751–754, 2016.

[232] Cancer Research UK. Skin cancer incidence statistics, 2017.

[233] Dirk Schadendorf and Axel Hauschild. Melanoma—the run of success continues. *Nature Reviews Clinical Oncology*, 11:75, 1 2014.

[234] Michael S Lawrence, Petar Stojanov, Paz Polak, Gregory V Kryukov, Kristian Cibulskis, Andrey Sivachenko, Scott L Carter, Chip Stewart, Craig H Mermel, Steven A Roberts, Adam Kiezun, Peter S Hammerman, Aaron McKenna, Yotam Drier, Lihua Zou, Alex H Ramos, Trevor J Pugh, Nicolas Stransky, Elena Helman, Jaegil Kim, Carrie Sougnez, Lauren Ambrogio, Elizabeth Nickerson, Erica Shefler, Maria L Cortés, Daniel Auclair, Gordon Saksena, Douglas Voet, Michael Noble, Daniel DiCara, Pei Lin, Lee Lichtenstein, David I Heiman, Timothy Fennell, Marcin Imielinski, Bryan Hernandez, Eran Hodis, Sylvan Baca, Austin M Dulak, Jens Lohr, Dan-Avi Landau, Catherine J Wu, Jorge Melendez-Zajgla, Alfredo Hidalgo-Miranda, Amnon Koren, Steven A McCarroll, Jaume Mora, Ryan S Lee, Brian Crompton, Robert Onofrio, Melissa Parkin, Wendy Winckler, Kristin Ardlie, Stacey B Gabriel, Charles W M Roberts, Jaclyn A Biegel, Kimberly Stegmaier, Adam J Bass, Levi A Garraway, Matthew Meyerson, Todd R Golub, Dmitry A Gordenin, Shamil Sunyaev, Eric S Lander, and Gad Getz. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, 499(7457):214–8, 2013.

[235] Charles M. Balch, Jeffrey E. Gershenwald, Seng Jaw Soong, John F. Thompson, Michael B. Atkins, David R. Byrd, Antonio C. Buzaid, Alistair J. Cochran, Daniel G. Coit, Shouluan Ding, Alexander M. Eggermont, Keith T. Flaherty,

Phyllis A. Gimotty, John M. Kirkwood, Kelly M. McMasters, Martin C. Mihm, Donald L. Morton, Merrick I. Ross, Arthur J. Sober, and Vernon K. Sondak. Final version of 2009 AJCC melanoma staging and classification. *Journal of Clinical Oncology*, 27(36):6199–6206, 2009.

[236] Jedd D. Wolchok, Vanna Chiarion-Sileni, Rene Gonzalez, Piotr Rutkowski, Jean-Jacques Grob, C. Lance Cowey, Christopher D. Lao, John Wagstaff, Dirk Schadendorf, Pier F. Ferrucci, Michael Smylie, Reinhard Dummer, Andrew Hill, David Hogg, John Haanen, Matteo S. Carlino, Oliver Bechter, Michele Maio, Ivan Marquez-Rodas, Massimo Guidoboni, Grant McArthur, Celeste Lebbé, Paolo A. Ascierto, Georgina V. Long, Jonathan Cebon, Jeffrey Sosman, Michael A. Postow, Margaret K. Callahan, Dana Walker, Linda Rollin, Rafia Bhore, F. Stephen Hodi, and James Larkin. Overall Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *New England Journal of Medicine*, page NEJMoa1709684, 2017.

[237] M.B. Amin, S.B. Edge, and F.L et al. Greene. *AJCC Cancer Staging Manual.* New York, NY, 8th ed. edition, 2017.

[238] Office for National Statistics. Cancer survival by stage at diagnosis for England. Technical report, 2016.

[239] Dirk Schadendorf, David E. Fisher, Claus Garbe, Jeffrey E. Gershenwald, Jean Jacques Grob, Allan Halpern, Meenhard Herlyn, Michael A. Marchetti, Grant McArthur, Antoni Ribas, Alexander Roesch, and Axel Hauschild. Melanoma. *Nature Reviews Disease Primers*, 1(April):15003, 2015.

[240] Linda Titus-Ernstoff, Ann E Perry, Steven K Spencer, Jennifer Gibson, Jiao Ding, Bernard Cole, and Marc S Ernstoff. Multiple primary melanoma: two-year results from a population-based study. *Archives of Dermatology*, 142(4):433–438, 4 2006.

[241] Cristiane Benvenuto-Andrade, Achiama Oseitutu, Anna Liza Agero, and Ashfaq A Marghoob. Cutaneous melanoma: Surveillance of patients for recurrence and new primary melanomas. *Dermatologic Therapy*, 18(6):423–435, 2005.

[242] Ulrike Leiter, Petra G Buettner, Thomas K Eigentler, Andrea Forschner, Friedegund Meier, and Claus Garbe. Is detection of melanoma metastasis during surveillance in an early phase of development associated with a survival benefit? *Melanoma Research*, 20(3):240–246, 6 2010.

[243] Andrew Bottomley, Corneel Coens, Stefan Suciu, Mario Santinami, Willem Kruit, Alessandro Testori, Jeremy Marsden, Cornelis Punt, François Salès, Martin Gore, Rona MacKie, Zvonko Kusic, Reinhard Dummer, Poulam Patel, Dirk Schadendorf, Alain Spatz, Ulrich Keilholz, and Alexander Eggermont. Adjuvant therapy with pegylated interferon alfa-2b versus observation in resected stage III melanoma: A phase III randomized controlled trial of health-related quality of life and symptoms by the European Organisation for Research and Treatment of Cancer . *Journal of Clinical Oncology*, 27(18):2916–2923, 2009.

[244] Alexander M.M. Eggermont, Vanna Chiarion-Sileni, Jean Jacques Grob, Reinhard Dummer, Jedd D. Wolchok, Henrik Schmidt, Omid Hamid, Caroline Robert, Paolo A. Ascierto, Jon M. Richards, Céleste Lebbé, Virginia Ferraresi, Michael Smylie, Jeffrey S. Weber, Michele Maio, Cyril Konto, Axel Hoos, Veerle de Pril, Ravichandra Karra Gurunath, Gaetan de Schaetzen, Stefan Suciu, and Alessandro Testori. Adjuvant ipilimumab versus placebo after complete resection of high-risk stage III melanoma (EORTC 18071): A randomised, double-blind, phase 3 trial. *The Lancet Oncology*, 16(5):522–530, 2015.

[245] Jeffrey Weber, Mario Mandala, Michele Del Vecchio, Helen J. Gogas, Ana M. Arance, C. Lance Cowey, Stéphane Dalle, Michael Schenker, Vanna Chiarion-Sileni, Ivan Marquez-Rodas, Jean-Jacques Grob, Marcus O. Butler, Mark R. Middleton, Michele Maio, Victoria Atkinson, Paola Queirolo, Rene Gonzalez, Ragini R. Kudchadkar, Michael Smylie, Nicolas Meyer, Laurent Mortier, Michael B. Atkins, Georgina V. Long, Shailender Bhatia, Celeste Lebbé, Piotr Rutkowski, Kenji Yokota, Naoya Yamazaki, Tae M. Kim, Veerle de Pril, Javier Sabater, Anila Qureshi, James Larkin, and Paolo A. Ascierto. Adjuvant Nivolumab versus Ipilimumab in Resected Stage III or IV Melanoma. *New England Journal of Medicine*, page NEJMoa1709030, 2017.

[246] Paul B. Chapman, Axel Hauschild, Caroline Robert, John B. Haanen, Paolo Ascierto, James Larkin, Reinhard Dummer, Claus Garbe, Alessandro Testori, Michele Maio, David Hogg, Paul Lorigan, Celeste Lebbe, Thomas Jouary, Dirk Schadendorf, Antoni Ribas, Steven J. O'Day, Jeffrey A. Sosman, John M. Kirkwood, Alexander M.M. Eggermont, Brigitte Dreno, Keith Nolop, Jiang Li, Betty Nelson, Jeannie Hou, Richard J. Lee, Keith T. Flaherty, and Grant A. McArthur. Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation. *New England Journal of Medicine*, 364(26):2507–2516, 2011.

[247] Antoni Ribas and Keith T Flaherty. BRAF targeted therapy changes the treatment paradigm in melanoma, 5 2011.

[248] Axel Hauschild, Jean Jacques Grob, Lev V Demidov, Thomas Jouary, Ralf Gutzmer, Michael Millward, Piotr Rutkowski, Christian U Blank, Wilson H. Miller, Eckhart Kaempgen, Salvador Martín-Algarra, Boguslawa Karaszewska, Cornelia Mauch, Vanna Chiarion-Sileni, Anne Marie Martin, Suzanne Swann, Patricia Haney, Beloo Mirakhur, Mary E Guckert, Vicki Goodman, and Paul B Chapman. Dabrafenib in BRAF-mutated metastatic melanoma: A multicentre, open-label, phase 3 randomised controlled trial. *The Lancet*, 380(9839):358–365, 7 2012.

[249] Jeffrey A Sosman, Kevin B Kim, Lynn Schuchter, Rene Gonzalez, Anna C Pavlick, Jeffrey S Weber, Grant A McArthur, Thomas E Hutson, Stergios J Moschos, Keith T Flaherty, Peter Hersey, Richard Kefford, Donald Lawrence, Igor Puzanov, Karl D Lewis, Ravi K Amaravadi, Bartosz Chmielowski, H Jeffrey Lawrence, Yu Shyr, Fei Ye, Jiang Li, Keith B Nolop, Richard J Lee, Andrew K Joe, and Antoni Ribas. Survival in BRAF V600–Mutant Advanced Melanoma Treated with Vemurafenib. *New England Journal of Medicine*, 366(8):707–714, 2 2012.

[250] Caroline Robert, Luc Thomas, Igor Bondarenko, Steven O'Day, Jeffrey Weber, Claus Garbe, Celeste Lebbe, Jean-François Baurain, Alessandro Testori, Jean-Jacques Grob, Neville Davidson, Jon Richards, Michele Maio, Axel Hauschild, Wilson H. Miller, Pere Gascon, Michal Lotem, Kaan Harmankaya, Ramy Ibrahim, Stephen Francis, Tai-Tsang Chen, Rachel Humphrey, Axel Hoos, and Jedd D Wolchok. Ipilimumab plus Dacarbazine for Previously Untreated Metastatic Melanoma. *New England Journal of Medicine*, 364(26):2517–2526, 6 2011.

[251] F Stephen Hodi, Steven J O'Day, David F McDermott, Robert W Weber, Jeffrey A Sosman, John B Haanen, Rene Gonzalez, Caroline Robert, Dirk Schadendorf, Jessica C Hassel, Wallace Akerley, Alfons J M van den Eertwegh, Jose Lutzky, Paul Lorigan, Julia M Vaubel, Gerald P Linette, David Hogg, Christian H Ottensmeier, Celeste Lebbé, Christian Peschel, Ian Quirt, Joseph I Clark, Jedd D Wolchok, Jeffrey S Weber, Jason Tian, Michael J Yellin, Geoffrey M Nichol, Axel Hoos, and Walter J Urba. Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. *New England Journal of Medicine*, 363(8):711–723, 6 2010.

[252] Prof Lesley Seymour, Canadian Cancer, Trials Group, Jan Bogaerts, Andrea Perrone, Translational Medicine, Robert Ford, Clinical Trials, Imaging Consulting, Belle Mead, Prof Lawrence H Schwartz, New York, Presbyterian Hospital, Prof Sumithra Mandrekar, Mayo Clinic, and Nancy U Lin. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *Lancet Oncol.*, 18(3):e143–e152, 2017.

[253] Yvonne M. Saenger and Jedd D. Wolchok. The heterogeneity of the kinetics of response to ipilimumab in metastatic melanoma: Patient cases. *Cancer Immunity*, 8(October 2007):1–7, 2008.

[254] Jedd D. Wolchok, Axel Hoos, Steven O'Day, Jeffrey S. Weber, Omid Hamid, Celeste Lebbé, Michele Maio, Michael Binder, Oliver Bohnsack, Geoffrey Nichol, Rachel Humphrey, and F. Stephen Hodi. Guidelines for the evaluation of immune therapy activity in solid tumors: Immune-related response criteria. *Clinical Cancer Research*, 15(23):7412–7420, 2009.

[255] Shanique R Palmer, Lori A Erickson, Ilia Ichetovkin, Daniel J Knauer, and Svetomir N Markovic. Circulating Serologic and Molecular Biomarkers in Malignant Melanoma. *Mayo Clinic Proceedings*, 86(10):981–990, 10 2011.

[256] H B Guo, B Stoffel-Wagner, T Bierwirth, J Mezger, and D. Klingmüller. Clinical significance of serum S100 in metastatic malignant melanoma. *European Journal of Cancer*, 31 A(11):1898–1902, 10 1995.

[257] R Dummer, A Hauschild, M Guggenheim, U Keilholz, and G Pentheroudakis. Cutaneous melanoma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, 23(SUPPL. 7):86–91, 10 2012.

[258] Gregory A. Chang, Jyothirmayee S. Tadepalli, Yongzhao Shao, Yilong Zhang, Sarah Weiss, Eric Robinson, Cindy Spittle, Manohar Furtado, Dawne N. Shelton,

George Karlin-Neumann, Anna Pavlick, Iman Osman, and David Polsky. Sensitivity of plasma BRAFmutant and NRASmutant cell-free DNA assays to detect metastatic melanoma in patients with low RECIST scores and non-RECIST disease progression. *Molecular Oncology*, 10(1):157–65, 2016.

[259] Stephen Q. Wong, Jeanette M. Raleigh, Jason Callahan, Ismael A. Vergara, Sarah Ftouni, Athena Hatzimihalis, Andrew J. Colebatch, Jason Li, Timothy Semple, Kenneth Doig, Christopher Mintoff, Devbarna Sinha, Paul Yeh, Maria Joao Silva, Kathryn Alsop, Heather Thorne, David D. Bowtell, David E. Gyorki, Gisela Mir Arnau, Carleen Cullinane, Damien Kee, Benjamin Brady, Fergal Kelleher, Mark A. Dawson, Anthony T. Papenfuss, Mark Shackleton, Rodney J. Hicks, Grant A. McArthur, Shahneen Sandhu, and Sarah-Jane Dawson. Circulating Tumor DNA Analysis and Functional Imaging Provide Complementary Approaches for Comprehensive Disease Monitoring in Metastatic Melanoma. *JCO Precision Oncology*, (1):1–14, 2017.

[260] Ashleigh C Mcevoy, Lydia Warburton, Zeyad Al-ogaili, Liesl Celliers, Leslie Calapre, Michelle R Pereira, Muhammad A Khattak, Tarek M Meniawy, Michael Millward, Melanie Ziman, and Elin S Gray. Correlation between circulating tumour DNA and metabolic tumour burden in metastatic melanoma patients. pages 1–8, 2018.

[261] The Cancer Genome Atlas Network. Genomic Classification of Cutaneous Melanoma. *Cell*, 161(7):1681–1696, 2015.

[262] Jenny H. Lee, Georgina V. Long, Alexander M. Menzies, Serigne Lo, Alexander Guminski, Kataraina Whitbourne, Michelle Peranec, Richard Scolyer, Richard F. Kefford, Helen Rizos, and Matteo S. Carlino. Association between circulating tumor DNA and pseudoprogression in patients with metastatic melanoma treated with anti–programmed cell death 1 antibodies. *JAMA Oncology*, 4(5):717–721, 2018.

[263] NICE. NICE guideline [NG14] - Melanoma: assessment and management, 2015.

[264] Alexander M M Eggermont, Vanna Chiarion-Sileni, Jean-Jacques Grob, Reinhard Dummer, Jedd D Wolchok, Henrik Schmidt, Omid Hamid, Caroline Robert, Paolo A Ascierto, Jon M Richards, Céleste Lebbé, Virginia Ferraresi, Michael Smylie, Jeffrey S Weber, Michele Maio, Lars Bastholt, Laurent Mortier, Luc Thomas, Saad Tahir, Axel Hauschild, Jessica C Hassel, F Stephen Hodi, Corina Taitt, Veerle de Pril, Gaetan de Schaetzen, Stefan Suciu, and Alessandro Testori. Prolonged Survival in Stage III Melanoma with Ipilimumab Adjuvant Therapy. *New England Journal of Medicine*, 375(19):1845–1855, 10 2016.

[265] Georgina V. Long, Axel Hauschild, Mario Santinami, Victoria Atkinson, Mario Mandalà, Vanna Chiarion-Sileni, James Larkin, Marta Nyakas, Caroline Dutriaux, Andrew Haydon, Caroline Robert, Laurent Mortier, Jacob Schachter, Dirk Schadendorf, Thierry Lesimple, Ruth Plummer, Ran Ji, Pingkuan Zhang, Bijoyesh Mookerjee, Jeff Legos, Richard Kefford, Reinhard Dummer, and John M. Kirkwood. Adjuvant Dabrafenib plus Trametinib in Stage III BRAF-Mutated Melanoma. *New England Journal of Medicine*, 377(19):1813–1823, 2017.

[266] R J Lee, G Gremel, A Marshall, K A Myers, N Fisher, J Dunn, N Dhomen, P G Corrie, M R Middleton, P Lorigan, Richard Marais, and Richard Marais. Circulating tumor DNA predicts survival in patients with resected high risk stage II/III melanoma. *Ann. Oncol.*, 29(2):490–496, 2018.

[267] Eleonor Olsson, Christof Winter, Anthony George, Yilun Chen, Jillian Howlin, Man-hung Eric Tang, Malin Dahlgren, Ralph Schulz, Dorthe Grabau, Danielle Van Westen, Mårten Fernö, Christian Ingvar, Carsten Rose, Pär-ola Bendahl, Lisa Rydén, Åke Borg, Sofia K Gruvberger-saal, Helena Jernström, and Lao H Saal. Serial monitoring of circulating tumor DNA in patients with primary breast cancer for detection of occult metastatic disease. *EMBO Molecular Medicine*, 7(8):1034–1047, 2015.

[268] Monya Baker. Digital PCR hits its stride. *Nature Methods*, 9:541, 5 2012.

[269] Tara C. Gangadhar, Samantha L. Savitch, Stephanie S. Yee, Wei Xu, Alexander C. Huang, Shannon Harmon, David B. Lieberman, Devon Soucier, Ryan Fan, Taylor A. Black, Jennifer J. D. Morrissette, Neeraj Salathia, Jill Waters, Shile Zhang, Jonathan Toung, Paul van Hummelen, Jian-Bing Fan, Xiaowei Xu, Ravi K. Amaravadi, Lynn M. Schuchter, Giorgos C. Karakousis, Wei-Ting Hwang, and Erica L. Carpenter. Feasibility of monitoring advanced melanoma patients using cell-free DNA from plasma. *Pigment Cell & Melanoma Research*, 31(1):73–81, 2018.

[270] Alan E Siroy, Genevieve M Boland, Denái R Milton, Jason Roszik, Silva Frankian, Jared Malke, Lauren Haydu, Victor G Prieto, Michael Tetzlaff, Doina Ivan, Wei-Lien Wang, Carlos Torres-Cabala, Jonathan Curry, Sinchita Roy-Chowdhuri, Russell Broaddus, Asif Rashid, John Stewart, Jeffrey E Gershenwald, Rodabe N Amaria, Sapna P Patel, Nicholas E Papadopoulos, Agop Bedikian, Wen-Jen Hwu, Patrick Hwu, Adi Diab, Scott E Woodman, Kenneth D Aldape, Rajyalakshmi Luthra, Keyur P Patel, Kenna R Shaw, Gordon B Mills, John Mendelsohn, Funda Meric-Bernstam, Kevin B Kim, Mark J Routbort, Alexander J Lazar, and Michael A Davies. Beyond BRAF(V600): clinical mutation panel testing by next-generation sequencing in advanced melanoma. *The Journal of Investigative Dermatology*, 135(2):508–15, 2015.

[271] Daniel Lai and Gavin Ha. HMMcopy : A package for bias-free copy number estimation and robust CNA detection in tumour samples from WGS HTS data, 2014.

[272] Leslie Calapre, Lydia Warburton, Michael Millward, Mel Ziman, and Elin S. Gray. Circulating tumour DNA (ctDNA) as a liquid biopsy for melanoma. *Cancer Letters*, 404:62–69, 2017.

[273] Hubing Shi, Willy Hugo, Xiangju Kong, Aayoung Hong, Richard C. Koya, Gatien Moriceau, Thinle Chodon, Rongqing Guo, Douglas B. Johnson, Kimberly B. Dahlman, Mark C. Kelley, Richard F. Kefford, Bartosz Chmielowski, John A. Glaspy, Jeffrey A. Sosman, Nicolas Van Baren, Georgina V. Long, Antoni Ribas, and Roger S. Lo. Acquired resistance and clonal evolution in melanoma during BRAF inhibitor therapy. *Cancer Discovery*, 4(1), 2014.

[274] Ignacio Varela, Patrick Tarpey, Keiran Raine, Dachuan Huang, Choon Kiat Ong, Helen Davies, David Jones, Meng-lay Lin, Jon Teague, Graham Bignell, Adam Butler, Juok Cho, Gillian L Dalgliesh, Danushka Galappaththige, Claire Hardy, Mingming Jia, Calli Latimer, King Wai Lau, John Marshall, Stuart Mclaren, Andrew Menzies, Laura Mudie, Lucy Stebbings, A David, L F a Wessels, Stephane Richard, Richard J Kahnoski, and John Anema. Exome sequencing identifies frequent mutation of the SWI / SNF complex gene PBRM1 in renal carcinoma. *Nature*, 469(7331):539–542, 2011.

[275] Dincer Goksuluk, Selcuk Korkmaz, and Gokmen Zararsiz. easyROC: a web-tool for ROC curve analysis (ver. 1.3), 2016.

[276] Florent Mouliere, Dineika Chandrananda, Anna M Piskorz, Elizabeth K Moore, James Morris, Lise Barlebo Ahlborn, Richard Mair, Teodora Goranova, Francesco Marass, Katrin Heider, Jonathan C M Wan, Anna Supernat, Irena Hudecova, Ioannis Gounaris, Susana Ros, Mercedes Jimenez-linan, Javier Garcia-corbacho, Keval Patel, Olga Østrup, Suzanne Murphy, Matthew D Eldridge, Davina Gale, Grant D Stewart, Johanna Burge, Wendy N Cooper, Michiel S Van Der Heijden, Charles E Massie, Colin Watts, Pippa Corrie, Simon Pacey, Kevin M Brindle, Richard D Baird, and Morten Mau-sørensen. Enhanced detection of circulating tumor DNA by fragment size analysis. *Science Transl. Med.*, 4921(November):1–14, 2018.

[277] Aadel A. Chaudhuri, Jacob J. Chabon, Alexander F. Lovejoy, Aaron M. Newman, Henning Stehr, Tej D. Azad, Michael S. Khodadoust, Mohammad Shahrokh Esfahani, Chih Long Liu, Li Zhou, Florian Scherer, David M. Kurtz, Carmen Say, Justin N. Carter, David J. Merriott, Jonathan C. Dudley, Michael S. Binkley, Leslie Modlin, Sukhmani K. Padda, Michael F. Gensheimer, Robert B. West, Joseph B. Shrager, Joel W. Neal, Heather A. Wakelee, Billy W. Loo, Ash A. Alizadeh, and Maximilian Diehn. Early detection of molecular residual disease in localized lung cancer by circulating tumor DNA profiling. *Cancer Discovery*, 7(12):1394–1403, 2017.

[278] Eric A. Klein, Earl Hubbell, Tara Maddala, Alex Aravanis, John F. Beausang, Darya Filippova, Samuel Gross, Arash Jamshidi, Kathryn Kurtzman, Ling Shen, Anton Valouev, Oliver Venn, Nan Zhang, David A. Smith, Robert Tibshirani Timothy Joseph Yeatman, Richard Thomas Williams, Anne-Renee Hartman, Michael Seiden, and Minetta C. Liu. Abstract 12021: Development of a comprehensive cell-free DNA (cfDNA) assay for early detection of multiple tumor types: The Circulating Cell-free Genome Atlas (CCGA) study. *J Clin Oncol*, 36:suppl; abstr 12021, 2018.

[279] Christopher Abbosh, Nicolai J. Birkbak, and Charles Swanton. Early stage NSCLC — challenges to implementing ctDNA-based screening and MRD detection. *Nature Reviews Clinical Oncology*, page 1, 7 2018.

[280] Farideh Z. Bischoff, Dianne X. Dang, Deborah Marquez-Do, Denisse Martinez, Cassandra Horne, Dorothy E. Lewis, and Joe Leigh Simpson. Detecting fetal DNA from dried maternal blood spots: Another step towards broad scale

non-invasive prenatal genetic screening and feasible testing. *Reproductive BioMedicine Online*, 6(3):349–351, 2003.

[281] Yali Xiong, Stacey Jeronis, Barbara Hoffman, Dan A. Liebermann, and Ossie Geifman-Holtzman. First trimester noninvasive fetal RHD genotyping using maternal dried blood spots. *Prenatal Diagnosis*, 37(4):311–317, 2017.

[282] Ji Yun Lee, Xu Qing, Wei Xiumin, Bai Yali, Sangah Chi, So Hyeon Bak, Ho Yun Lee, Jong-Mu Sun, Se-Hoon Lee, Jin Seok Ahn, Eun Kyung Cho, Dong-Wan Kim, Hye Ryun Kim, Young Joo Min, Sin-Ho Jung, Keunchil Park, Mao Mao, and Myung-Ju Ahn. Longitudinal monitoring of EGFR mutations in plasma predicts outcomes of NSCLC patients treated with EGFR TKIs: Korean Lung Cancer Consortium (KLCC-12-02). *Oncotarget*, 7(6):6984–6993, 2016.

[283] Rubicon Genomics. Targeted Capture of ThruPLEX® Libraries with Agilent SureSelect®XT Target Enrichment System.

[284] Maura Costello, Trevor J. Pugh, Timothy J. Fennell, Chip Stewart, Lee Lichtenstein, James C. Meldrim, Jennifer L. Fostel, Dennis C. Friedrich, Danielle Perrin, Danielle Dionne, Sharon Kim, Stacey B. Gabriel, Eric S. Lander, Sheila Fisher, and Gad Getz. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Research*, 41(6):1–12, 2013.

[285] University of Michigan. Connor - METHODS, 2016.

[286] Jiajie Zhang, Kassian Kobert, Tomáš Flouri, and Alexandros Stamatakis. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, 30(5):614–620, 2014.

[287] Mónica López-Ratón, María Xosé Rodríguez-Álvarez, Carmen Cadarso Suárez, and Francisco Gude Sampedro. OptimalCutpoints : An R Package for Selecting Optimal Cutpoints in Diagnostic Tests. *Journal of Statistical Software*, 61(8):1–36, 2014.

[288] Daniela Sint, Lorna Raso, and Michael Traugott. Advances in multiplex PCR: Balancing primer efficiencies and improving detection success. *Methods in Ecology and Evolution*, 3(5):898–905, 2012.

[289] Kristian Cibulskis, Michael S Lawrence, Scott L Carter, Andrey Sivachenko, David Jaffe, Carrie Sougnez, Stacey Gabriel, Matthew Meyerson, Eric S Lander, and Gad Getz. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology*, 31(3):213–219, 3 2013.

[290] Olena Kis, Rayan Kaedbey, Signy Chow, Arnavaz Danesh, Mark Dowar, Tiantian Li, Zhihua Li, Jessica Liu, Mark Mansour, Esther Masih-Khan, Tong Zhang, Scott V Bratman, Amit M Oza, Suzanne Kamel-Reid, Suzanne Trudel, and Trevor J Pugh. Circulating tumour DNA sequence analysis as an alternative to multiple myeloma bone marrow aspirates. *Nature Communications*, 8:15086, 5 2017.

[291] Joshua D Cohen, Lu Li, Yuxuan Wang, Christopher Thoburn, Bahman Afsari, Ludmila Danilova, Christopher Douville, Ammar A Javed, Fay Wong, Austin Mattox, Ralph. H Hruban, Christopher L Wolfgang, Michael G Goggins, Marco Dal Molin, Tian-Li Wang, Richard Roden, Alison P Klein, Janine Ptak, Lisa Dobbyn, Joy Schaefer, Natalie Silliman, Maria Popoli, Joshua T Vogelstein, James D Browne, Robert E Schoen, Randall E Brand, Jeanne Tie, Peter Gibbs, Hui-Li Wong, Aaron S Mansfield, Jin Jen, Samir M Hanash, Massimo Falconi, Peter J Allen, Shibin Zhou, Chetan Bettegowda, Luis A Diaz, Cristian Tomasetti, Kenneth W Kinzler, Bert Vogelstein, Anne Marie Lennon, and Nickolas Papadopoulos. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science*, 1 2018.

[292] Sumonta Chaisomchit, Rattanawadee Wichajarn, Noppavan Janejai, and Wiyada Chareonsiriwatana. Stability of genomic DNA in dried blood spots stored on filter paper. *Southeast Asian Journal of Tropical Medicine and Public Health*, 36(1):270–273, 1 2005.

[293] Timothy M Blicharz, Ping Gong, Bernard M Bunner, Larry L Chu, Kaela M Leonard, Jessica A Wakefield, Richard E Williams, Maisam Dadgar, Carlo A Tagliabue, Ragheb El Khaja, Stephanie L Marlin, Ramin Haghgooie, Shawn P Davis, Donald E Chickering, and Howard Bernstein. Microneedle-based device for the one-step painless collection of capillary blood samples. *Nature Biomedical Engineering*, 2(3):151–157, 2018.

[294] BusinessWire. FDA Approves Foundation Medicine's FoundationOne CDx™, the First and Only Comprehensive Genomic Profiling Test for All Solid Tumors Incorporating Multiple Companion Diagnostics, 2017.

[295] Ian Sample. Routine DNA tests will put NHS at the 'forefront of medicine', 2018.

[296] Clare Turnbull, Richard H. Scott, Ellen Thomas, Louise Jones, Nirupa Murugaesu, Freya Boardman Pretty, Dina Halai, Emma Baple, Clare Craig, Angela Hamblin, Shirley Henderson, Christine Patch, Amanda O'Neill, Andrew Devereaux, Katherine Smith, Antonio Rueda Martin, Alona Sosinsky, Ellen M. McDonagh, Razvan Sultana, Michael Mueller, Damian Smedley, Adam Toms, Lisa Dinh, Tom Fowler, Mark Bale, Tim Hubbard, Augusto Rendon, Sue Hill, and Mark J. Caulfield. The 100 000 Genomes Project: Bringing whole genome sequencing to the NHS. *BMJ*, 361(April):1–7, 2018.